

OM of: US-09-768-877-2 to: GenBank.* out_format: pfs

Date: Sep 17, 2002 6:32 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-MODEL-firmet-p2n.model -DEV-xip
-O/-qg2.1/USPTO.spool/US09768877/rnalt.16092002.152811.8346/app_query.fasta.1.739
-DB-GenBank -OPMT-fastap -SUFFIX-ige -GAPOP-12.000 -GAPEXT-4.000
-KMA-MATCH-0.100 -LOOPCL-0.000 -LOOPEXT-0.000 -GAPOP-4.500
-OGAPEXT-0.050 -XGAPOP-10.000 -XGAPEXT-0.500 -FGAPOP-6.000
-FAAPEXT-7.000 -YGAPOP-10.000 -YGAPEXT-0.500 -DELOP-6.000
-DELEXT-7.000 -START-1 -MATRIX-biosum62 -TRANS-human40.cdi
-LIST-45 -DOCALLIGN-200 -THR SCORE-pct -THR MAX-100 -THR MIN-0
-ALIGN-15 -MODE-LOCAL -OUTPMT-pfs -NORM-ext -HEAPSIZ-500
-MINLEN-0 -MAXLEN-200000000 -USER-US09768877.ecgml.1.8402
-NCPU-6 -ICPU-3 -LONGLOG -DEV-TIMEOUT-120 -WARN-TIMEOUT-30
-NO_XLPEX -WAIT -THREADS-1
```

Search information block:

```
Query: US-09-768-877-2
Query length: 672
Database: GenBank.*
Database sequences: 1797656
Database length: 187333701
Search time (sec): 2930.000000
```

score_list:

Sequence	Strd Orig	ZScore	EScore	len	Document
gb.pr:AB058748	+ 3649.00	3089.26	1.0e-163	2536	AB058748 Homo sapiens mRNA for
gb.pr:BC004260	+ 3649.00	3089.00	1.1e-163	2614	BC004260 Homo sapiens, calpain
gb.pr:BC007553	+ 3649.00	3089.00	1.1e-163	2614	BC007553 Homo sapiens, clone
gb.pr:AR153290	+ 3649.00	3088.98	1.1e-163	2620	AR153290 Sequence 3 from paten
gb.pr:AF089088	+ 3649.00	3088.98	1.1e-163	2620	AF089088 Homo sapiens calpain
gb.pr:AR153295	+ 3604.00	3051.46	1.3e-161	2455	AR153295 Sequence 13 from paten
gb.pr:AF089094	+ 3604.00	3051.46	1.3e-161	2455	AF089094 Homo sapiens calpain
gb.pr:AB072744	+ 3388.50	2868.60	2.0e-151	2602	AB072744 Macaca fascicularis t
gb.pr:AR153291	+ 3290.00	2786.32	7.7e-147	2297	AR153291 Sequence 5 from paten
gb.pr:AF089090	+ 3290.00	2786.32	7.7e-147	2297	AF089090 Homo sapiens calpain
gb.pr:AR153294	+ 3239.50	2742.80	2.1e-144	2516	AR153294 Sequence 11 from paten
gb.pr:AF089093	+ 3239.50	2742.80	2.1e-144	2516	AF089093 Homo sapiens calpain
gb.pr:AR153293	+ 3090.50	2617.85	1.9e-137	2204	AR153293 Sequence 9 from paten
gb.pr:AF089092	+ 3090.50	2617.85	1.9e-137	2204	AF089092 Homo sapiens calpain
gb.pr:BC010969	+ 2969.00	2513.85	1.2e-131	2528	BC010969 Mus musculus, calpain
gb.pr:BC005681	+ 2969.00	2513.85	1.2e-131	2528	BC005681 Mus musculus, calpain
gb.pr:AR153298	+ 2963.00	2508.83	2.2e-131	2511	AR153298 Sequence 19 from paten
gb.pr:AF089089	+ 2963.00	2507.59	2.6e-131	2901	AF089089 Mus musculus calpain
gb.pr:AF089099	+ 2958.00	2505.09	3.6e-131	2372	AF089099 Rattus norvegicus cal
gb.pr:AF126867	+ 2941.00	2490.45	2.3e-130	2442	AF126867 Mus musculus calpain
gb.pr:AR153198	+ 2726.50	2312.89	1.8e-120	1542	AR153198 Sequence 13 from paten
gb.pr:AR153610	+ 2715.50	2306.90	3.9e-120	1534	AR153610 Sequence 3 from paten
gb.pr:AR153608	+ 2715.50	2303.99	5.7e-120	2180	AR153608 Sequence 1 from paten
gb.pr:AR153292	+ 2715.50	2301.34	8.0e-120	2001	AR153292 Sequence 7 from paten
gb.pr:AF089091	+ 2715.50	2301.34	8.0e-120	2001	AF089091 Homo sapiens calpain
gb.pr:AK027330	+ 2487.50	2110.68	3.3e-109	1534	AK027330 Homo sapiens cDNA FL
gb.pr:AC080022	+ 2043.00	1693.13	6.0e-86	17197	AC080022 Homo sapiens chrom
gb.pr:AF158748	+ 2015.50	1678.98	3.7e-85	6565	AF158748 Homo sapiens map 243
gb.pr:AR153289	+ 1989.00	1659.05	4.8e-84	49136	AR153289 Sequence 1 from paten
gb.pr:AF203031	+ 1294.50	1077.97	1.1e-51	22797	AF203031 Mus musculus calpain
gb.pr:AR153014	+ 665.00	561.24	4.6e-23	2498	AR153014 Sequence 5 from paten
gb.pr:AF263610	+ 665.00	561.24	4.6e-23	2498	AF263610 Sus scrofa microsomal
gb.pr:AF263610	+ 655.50	553.80	1.7e-22	2994	AF263610 Sus scrofa microsomal
gb.pr:AB007775	+ 652.50	569.32	2.4e-23	468	AB007775 Sequence 2944 from pat
gb.pr:AF263609	+ 649.00	548.15	3.4e-22	2639	AF263609 Sus scrofa microsomal
gb.pr:AF263609	+ 649.00	548.15	3.4e-22	2639	AF263609 Sus scrofa microsomal
gb.pr:AF263609	+ 644.50	548.54	3.4e-22	2061	AF263609 Sequence 7 from paten
gb.pr:AX090201	+ 644.50	548.54	3.4e-22	2109	AX090201 Sequence 5 from paten
gb.pr:AX090209	+ 644.50	548.34	3.5e-22	2109	AX090209 Sequence 12 from paten

```
seq_name: gb.pr:AB058748
gb.pr:AX090207 + 644.50 548.09 3.6e-22 2172 | AX090207 Sequence 3 from
gb.pr:AX090205 + 644.50 547.90 3.7e-22 2220 | AX090205 Sequence 1 from
gb.pr:AX063615 + 644.50 547.46 3.9e-22 2338 | AX063615 Sequence 1 from
gb.pr:HS242832 + 644.50 547.46 3.9e-22 2338 | A1242832 Homo sapiens m
gb.pr:AX090213 + 644.50 545.89 4.8e-22 2806 | AX090213 Sequence 9 from
```

seq_documentation block:

```
LOCUS AB058748 2536 bp mRNA linear PRI 05-JUN-2001
DEFINITION Homo sapiens mRNA for KIAA1845 protein, partial cds.
ACCESSION AB058748
VERSION AB058748.1 GI:14017906
KEYWORDS
SOURCE
ORGANISM Homo sapiens brain cDNA to mRNA, clone:f905317.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
```

REFERENCE

```
1 (sites) Nagase,T., Nakayama,M., Nakajima,D., Kikuno,R. and Ohara,O.
Prediction of the coding sequences of unidentified human genes. XX.
The complete sequences of 100 new cDNA clones from brain which code
for large proteins in vitro
DNA Res. 8 (2), 85-95 (2001)
2 (bases 1 to 2536)
21245130
Ohara,O., Nagase,T. and Kikuno,R.
Direct Submission
Submitted (27-MAR-2001) Osamu Ohara, Kazusa DNA Research Institute,
Department of Human Gene Research; 1532-3, Yana, Kisarazu, Chiba
292-0812, Japan (E-mail:cdna@info.kazusa.or.jp,
URL:http://www.kazusa.or.jp/huge, Tel:81-438-52-3913,
Fax:81-438-52-3914)
```

FEATURES

```
source
1..2536
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="f905317"
/tissue-type="brain"
/note="vector:pluascriptII SK plus"
1..2119
/gene="KIAA1845"
<1..2119
/gene="KIAA1845"
/note="Start codon is not identified, f905317 cDNA clone
for KIAA1845 has 921-bp, 1755-bp, 495-bp and 1014-bp
insertions after the positions 570, 788, 930 and 1378 of
the sequence of KIAA1845, respectively."
/codon_start=2
/product="KIAA1845 protein"
/protein_id="F8A474.1"
/db_xref="GI:14017907"
/translation="GGGGEPAADSPSLGCDPEATGCRMEPAEPRAAGCATPAR
ELFRDAFPAADSLFCDLSTPLAARDITMRPEICATGATFDDREGVKGGL
LGCWFLCAALAKSRHLLDVIIPGSGMADYRGSEFTORIMQFGWVEYTDRL
LPDIAGRLFCRQREDVDFMLPELVKACGHSYDRAAGVADLDTGLIARW
NKGVAAGSGGQDRPWRHRTCTRLHKKDCICISCVLSPAGARELGEHFAIVS
DARELOGAGGQCLILRTONPWGRKMGIMRGSGMSQVAAVASLLOQGEF
WWEERELFEPELTVGYPTENGHLQSLYTERLLCHRTALPGAWKQSGAGCRNRS
GPGSNPKFWLRVSEPEVYIAVQSRSLAAADWAGRAALVDSHTSNGPASIPKHY
QAVGLHLMKVEKRVNLPVLSMPVAGACAAHVDREVLRCGLSPGYLAIVSPFLK
DAGEFLVFTSGTRVLSAIVAKNTPTGAAIADGEGYVQLGSRVAGGTAGGSR
NPAATYDNPNCSPSPVGGGRCVCTTTLQHKRPSDTEHPHGFHIVDPEGRCODA
PPLDPEPLSLCPHRTADEVSRCLLPAGTYKVVPESTYLPDTGAFVYTIATRIDRP
SIHSEMLGQFLOEVSVAVMYK"
```

BASE COUNT

```
438 a 795 c 862 g 441 t
```

alignment_scores:

```
Quality: 3649.00 Length: 672
Ratio: 5.430 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
```

alignment_block:
US-09-768-877-2 x AB058748 ..
Align seg 1/1 to: AB058748 from: 1 to: 2536

1 MetArgAlaGlyArgGlyAlaThrProAlaArgLeuPheArgAspAl 17
101 ATCGGGCGGGCGGGCGGCGGAGCCGCGGAGGAGCTGTTCGGGAGCG 150

17 aAlaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrProL 34
151 CGCCTTCCCGCGCGGAGCTCTCGCTCTCTCTCGGACTGTCTACGCCGC 200

34 euAlaGlnPheArgGluAspIleThrTrpArgArgProGlnGluIleCys 50
201 TGGCCCAAGTTCCCGGAGGACATCACGTGGAGCGGCCCCAGGAGATTGT 250

51 AlaThrProArgLeuPheProAspAspProArgGluGlnVallysGI 67
251 GCCACACCCCGGCTGTTTCCAGATGACCCACGGGAAGGGGAGGTGAAGCA 300

67 nGlyLeuLeuGlyAspCysTrpPheLeuCysAlaCysAlaAlaLeuGlnL 84
301 GGGGCTGCTGGGGGATGTGGTTCCTGTGTGCTGGCGCGCGCTGCAGA 350

84 ysSerArgHisLeuLeuAspGlnValIleProProGlyGlnProSerTrp 100
351 AGAGCAGGACCTCTCTGGACAGGTCACTTCTCCGGGACACGCCAGCTGG 400

101 AlaAspGlnGluThrArgGlySerPheThrCysArgIleTrpGlnPheGI 117
401 GCGGACAGAGTACCGGGGCTCTTCCACCTGTGCGCATTTGGCAGTTTG 450

117 yArgTrpValGluValThrThrAspAspArgLeuProCysLeuAlaGlyA 134
451 ACGCTGGGTGGAGTGACACACAGATACCGGCTGCCGTGCTTGCAGGGA 500

134 rgLeuCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
501 GACTCTGTTTCTCCCGCTGCCAGAGGAGGATGTGTTCTGGCTCCGCTTA 550

151 LeuGluLysValTyrrAlaLysValHisGlySerTyrrGluHisLeuTrpAl 167
551 CTGGAAGAGGTCTACGCCAAGGTCCATGGGTCTACGACACCTGTGGGC 600

167 agGlyGlnValAlaAspAlaLeuValAspLeuThrGlyGlyLeuAlaGluA 184
601 CGGGCAGGTGGCGGATGCCCTGGTGGACCTTGACCGCGCGCTGGCAGAAA 650

184 rgTrpAsnLeuLysGlyValAlaGlySerGlyGlnGlnAspArgPro 200
651 GATGGAACCTGAAGGCGGTAGCAGGAAGCGGAGCGGAGGAGGAGGCA 700

201 GlyArgTrpGluHisArgThrCysArgGlnLeuLeuHisLeuLysAspGI 217
701 GCGCGCTGGAGCACAGGACTGTGCGGAGCTGTCCACCTGAAGGAGCA 750

217 nCysLeuIleSerCysValLeuSerProArgAlaGlyAlaArgGluL 234
751 GTGCTGTGATCAGTGTGCGTGTCTAGCCAGGAGCAGAGTGTCCCGGAGC 800

234 euGlyCluPheHisAlaPheIleValSerAspLeuArgGluLeuGlnGly 250
801 TGGGGAGGTTCATGCTTTCATTTGTCTCGGACCTCGGGAGGTCCAGGGT 850

251 GlnAlaGlyGlnCysIleLeuLeuLeuArgIleGlnAsnProTrpGlyAr 267
851 CAGGCGGGCCAGTGCATCTGCTGTGCGGATCCAGAACCCCTGGGGCCG 900

267 gArgCysTrpGlnGlyLeuTrpArgGluGlyGlyGlyTrpSerGlnV 284
901 GCGGTGCTGGCAGGGGCTCTGGAGAGAGGGGGGTGAAGGCTGGAGCCAGG 950

284 aAlaPalaAlaValAlaSerGluLeuLeuSerGlnLeuGlnGluGlyGlu 300
951 TAGATGCGAGCGGTAGCATCTGAGCTCTCTGCCAGCTCCAGGAGGGGAG 1000

301 PheTrpValGluGluGluGluPheLeuArgGluPheAspGluLeuThrVa 317
1001 TTCTGGGTGGAGGAGGAGGAGTTCTTCAGGAGTTTTCAGCAGCTCACCGT 1050

317 lGlyTyrrProValThrGluAlaGlyHisLeuGlnSerLeuTyrrThrGluA 334
1051 TGCTACCCCGGTACGGAGGCGGCCACCTGCAGAGCCTCTACACAGAGA 1100

334 rgLeuLeuCysHisThrArgAlaLeuProGlyAlaTrpValLysGlyGln 350
1101 GGCTGCTCTGCCATACGCGGCGCTGCTGGGCGCTGGGTCAAGGGCCAG 1150

351 SerAlaGlyGlyCysArgAsnAsnSerSerGlyPheProSerAsnProLysPh 367
1151 TCAGCAGGAGGCTGCCGGAAACACAGCGGCTTTCACAGCAACCCCAATT 1200

367 eTrpLeuArgValSerGluProSerGluValTyrrIleAlaValLeuGlnA 384
1201 CTGGCTGGCGGTCTCAGAACCGAGTGAGTGTACATTGGCGCTCTGCAGA 1250

384 rgSerArgLeuHisAlaAlaAspTrpAlaGlyArgAlaArgAlaLeuVal 400
1251 GATCCAGGCTGCACGCGCGGACTGGGCGCGCGCGCGGCGGACACGTGGTG 1300

401 GlyAspSerHisThrSerTrpSerProAlaSerIleProGlyLysHisTy 417
1301 GGTGACAGTCATACTTCTGGAGCCAGCGAGCATCCCGGGCAAGCACTA 1350

417 rGlnAlaValGlyLeuHisTrpLysValGluLysArgArgValAsnL 434
1351 CCAGGCTGGGTCTGCACCTCTGGAGGTAGAGAGCGCGGGTCAATC 1400

434 euProArgValLeuSerMetProProValAlaGlyThrAlaCysHisAla 450
1401 TGCTAGGCTCTGTCTCATGCCCCCGTGGTGGCACCGGCTCCCATGCA 1450

451 TyrAspArgGluValHisLeuArgCysGluLeuSerProGlyTyrrTyrl 467
1451 TAGGACCGGGAGGTCCACCTGTGTGAGCTCTCACCGGGCTACTACCT 1500

467 uAlaValProSerThrPheLeuLysAspAlaProGlyGluPheLeuA 484
1501 GGCTGTCCCCCAGCACCTTCTGNAGGACGCGCGAGGGGAGTTCTCTCTCC 1550

484 rgValPheSerThrGlyArgValSerLeuSerAlaIleArgAlaValAla 500
1551 GAGTCTTCTTACCGGGCGAGTCTCCCTTAGCCCATCAGGCGAGTGGCC 1600

501 LysAsnThrThrProGlyAlaAlaLeuProAlaGlyGluTrpGlyThrVa 517
1601 AAGAACACACCCCGGGGAGCCTTCTGCGGGGAGTGGGGGAGCCCT 1650

517 lGlnLeuArgGlySerTrpArgValGlyGlnThrAlaGlyGlySerArgA 534
1651 GCAGCTACGGGGTCTTGGAGAGTGGCCGAGCGCGGGGGGAGCAGAGA 1700

534 snPheAlaSerTyrrProThrAsnProCysPheProPheSerValProGlu 550
1701 ACTTTGGCTCATACCCCAACCCCTGCTTCCCTTCTCTCGGCTCCCGAG 1750

551 GlyProGlyProArgCysValArgIleThrLeuHisGlnHisCysArgPr 567
1751 GGCCCTGGCCCCCGCTGGCTCCGATCATCTGTCATCAGCAGCTGCGGCC 1800

567 oSerAspThrGluPheHisProIleGlyPheHisIlePheGlnValProG 584
1801 CAGTGACACCGAGTTCACCCCATCGGCTTCCATATCTTCCAGGTCCAG 1850

us-09-768-877-2.rge

```

584  IuGIyGLyArVgSerGlnaSpAlaProrProLeuLeuGlnGluProLeu 600
1851 AGGGTGGAAAGAGCCAGGAGCCACCCACCCTGCTGCTGCAGGAGCCGCTG 1900
601  LeuSerGysValProHisArgTYrAlaGlnGluValSerArgLeuGlyLe 617
1901 CTGAGCTGCCTGCCACATCCCTTACGCCCAAGAGGTGAGCGGCTTCCT 1950
617  uLeuProAlaGlyThrTYrLySValValProSerThrTYrLeuProAsp 634
1951 CTGCGCTGCAGGACACCTACAAAGTTGTGGCCCTCCACCTACTGCGGACA 2000
634  hrcLnuGlyAlaPheThrValThrIleAlaThrArgIleAspArgProSer 650
2001 CAAAGGGGGGCTTCACAGTGCATCCATCGCAACCAAGATTGACAGGCCATCC 2050
651  IleHisSerGlnGluMetLeuGlyGlnPheLeuGlnGluValSerValMe 667
2051 ATTCAAGCCAGGAGATGCTGGGCGCAGTTCCTCCAAAGAGTCTCCGCGCAT 2100
667  talAlaValMetIysThr 672
2101 GGCAGTGAATGAANAACC 2116
seq_name: gb_pr:BC004260

```

```

/cclone="MGC:10770 IMAGE:3606959"
/clone_type="uterus, endometrium adenocarcinoma"
/clone_lib="NH_MGC_4"
/lab_host="DH10B-R"
/note="vector: pOTB7"
146. . 2164
/codon_start=1
/product="calpain 10"
/protein_id="AA04260.1"
/db_xref="GI:13279050"
/translaton="MARGATPARELLEFDAAPFADSLFCDLSTPLAOFREDITWR
RPEICLPTPLFPDDPRAGVQKGLGIDCWFELCALGSRHLHLOVITPGQSPAD
QERKSGFTCRIMQGRMVEVTTDDRLPLCLCFSGCSQEDVPTMLEKVTAKHG
SYELMAGQVADALVDTGLGLERBMNKLKAGVSGGODDGRGRRHETROLHLKDDC
LISCSVSPNAGARELGEHFATFVSDIRELOGAGOCILLRLQNHQRCMGOIMEE
GGEBSOVDAVASSELISQDEGEFWEEEEFLEPDELTVGPVPEAGHLSQSLYTER
LICHTRALPGAWKQSGAGGCRNNSGPDKFMLERVSESEVYIAVLQSRHLAAM
AGARARLVCDSLTSWSPASIPGKHYOAGVGLMKVEKRRNRLPRVIMSPVAGTACA
YDEVEHLRCLSPGYTLAVSTFLKAPGELLRLRVSTGVSISATIRAVAKNTTPGAA
LIPGEGEYVOLRGSWMRGQAGGSRNDSYPTNPCCFPSTVEGPGRCVAVITLHOB
RSPTEHFGICFHTFQVPEGRSGDAPPLAQEPLVLCVPRVIAQEVSRCLCLPAGTVK
VYSTLIPDEGAFVYIATRIDRPSIHSQMLGQFQEVSVAAVMKT"

```

alignment scores:

Quality:	3649.00	Length:	672
Ratio:	5.430	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	100.000

alignment_block:

Align seg 1/1 to: BC004260 from: 1 to: 2614

I MetArgAlaGlyArgGlyAlaThrProAlaArgGluLeuPheArgGAspAl 17

[illegible]

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

34 [auai.ac.jp/hokuzetsu/](#) [http://www.earth.ri.ac.jp/](#)

246 TGGCCAGTTC CGCGAGGACATCAGTGGAAGCGGCCAACCAATTCT

51 AlaThrProArgLeuPheProAspAspProArgGluGlnValLysG1 67

236 GGCACACCCGGCGTTCAGATGACCCACGGGAGGGCAGGTGAACA 345

84

2

[illegible]

101 AlaAspGlnGluTyrArgGlySerPheThrCysArgIleTrnGlnPheG1 117

446 GCCGACCAGGAGTACCGGGGCTCCTTCACCTGTCGCAATTGGCAGTTTGG 495

11 ymgiipvalgiuvalitititraspaspargleuprocysleuagllya 134

.....GCGCCGCGCAGGGG 345

EAC

151 laug)llvsva|tvra|atwsva|hieclwGortmclw|atwvva|lcc

|||||
596 CTGAAAGGTCTACGCCAGAGGTCCATGGGTCTCTACAGCACCTGTGGGC 645
|||||
167 aclyGlnValAlaAspAlaLeuValAspLeuThrGlyGlyLeuAlaGluA 184
|||||
646 CGGCAGGTGGCGATGCCCTGGTGGACCTTGACCGCGCCCTGGCAGAA 695
|||||
184 rGTPAsnLeuLysGlyValAlaGlySerGlyGlyGlnGlnAspArgPro 200
|||||
696 GATGAAACCTGAAGGGCGTAGCAGGAAGCGAGGCCACAGCAGGACGCCA 745
|||||
201 GlyArgTrpGluHisArgThrCysArgGlnLeuLeuHisLeuLysaspG 217
|||||
746 GGCCGCTGGGAGCACAGGACTTCTGGCAGCTGCTCCACCTGAAGGACCA 795
|||||
217 nCysLeuIleSerCysCysValLeuSerProArgAlaGlyAlaArgGluL 234
|||||
796 GTGTCTGATCAGCTGCTGCTGCTCAGCCACAGCAGAGGTGCCGGAGC 845
|||||
234 euGlyGluPheHisAlaPheIleValSerAspLeuArgGluLeuGlnGly 250
|||||
846 TGGGGAGGTCCATGCCCTTCTTCGACCTCGGAGCTCGGGAGCTCCAGGT 895
|||||
251 GlnAlaGlyGlnCysIleLeuLeuLeuLeuArgIleGlnAsnProTrpGlyAr 267
|||||
896 CAGCGGGCCAGTGCAATCTCTGCTGCTGCGGATCCAGAACCCCTGGGGCCG 945
|||||
267 gArGysTrpGlnGlyLeuTrpArgGluGlyGlyGluGlyTrpSerGlnV 284
|||||
946 CGGTGCTGGCAGGGGCTCTGGAGAGAGGGGGGTGAAGGTGGAGCCAGG 995
|||||
284 alAspAlaValAlaSerGluLeuLeuSerGlnLeuGlnGluGlyGlu 300
|||||
996 TAGATGACGGGTAGCATCTGAGCTCTCTGCCAGCTCCAGGAAGGGAG 1045
|||||
301 PheTrpValGluGluGluPheLeuAtqGluPheAspGluLeuThrVa 317
|||||
1046 TTCGGTGGAGGAGGAGGAGTTCCTCAGGGAGTTTGACAGGCTCACCGT 1095
|||||
317 lGlyTrpValThrGluAlaGlyHisLeuGlnSerLeuTrpThrGluA 334
|||||
1096 TGGCTACCCGGTCACGGAGGCGGCCACCTGCAGAGCCTTACACAGAGA 1145
|||||
334 rGleuLeuCysHisThrArgAlaLeuProGlyAlaTrpValLysGlyGln 350
|||||
1146 GGTGCTCTCCATACGGCGGGCGCTGCTGGGGCCCTGGGTCAAGGGCCAG 1195
|||||
351 SerAlaGlyCysArgAsnAsnSerGlyPheProSerAsnProLysph 367
|||||
1196 TCAGCAGGAGGCTGCCGGAACACAGCGGCTTTCCAGCAACCCCAAAAT 1245
|||||
367 eTrpLeuArgValSerGluProSerGluValTyrlleAlaValLeuGlnA 384
|||||
1246 CTGGCTCGGGTCTCAGAACCGAGTCAGGTGTACATTGGCGTCTGCAGA 1295
|||||
384 rGSerArgLeuHisAlaAlaAspTrpAlaGlyArgAlaArgAlaLeuVal 400
|||||
1296 GATCCAGGCTGCACGGCGGAGTGGGCAGCGCGCGCGGCGGCTGGTG 1345
|||||
401 GlyAspSerHisThrSerTrpSerProAlaSerIleProGlyLysHisTy 417
|||||
1346 GGTGACAGCTATCTTCTGGAGGCCAGCGAGCATCCCGGGCAAGCACTA 1395
|||||
417 rGlnAlaValGlyLeuHisLeuTrpLysValGluLysArgArgValAsnL 434
|||||
1396 CCAGGCTGTGGGTCTGCACCTCTGGAAGGTAGAGAGCGGGGTCAATC 1445
|||||
434 euProArgValLeuSerMetProProValAlaGlyThrAlaCysHisAla 450
|||||
1446 TGCTAGGGTCTCTCATGCCCCCGCTGGCTGGCACCGCTGCCATGCA 1495
|||||
451 TyrAspArgGluValHisLeuArgCysGluLeuSerProGlyTyrlle 467
|||||

1496 TAGACCGGGAGGTCCACCTCGCTTGTGAGCTCTCACCGGGCTACTACCT 1545
|||||
467 uAlaValProSerThrPheLeuLysaspAlaProGlyGluPheLeuLeuA 484
|||||
1546 GGCTGTCCCGACGACCTTCTGAAGGACGCCAGGGGAGTTCTCGCTCC 1595
|||||
484 rGValPheSerThrGlyArgValSerLeuSerAlaIleArgAlaValAla 500
|||||
1596 GAGTCTTCTACCGGGGAGTCTCCCTTAGCGCCATCAGGGCAGTGGCC 1645
|||||
501 LysAsnThrThrProGlyAlaAlaLeuProAlaGlyGluTrpGlyThrVa 517
|||||
1646 AAGAACACCAACCCCGGGGAGCCCTCGCTGCGGGGAGTGGGGACCGT 1695
|||||
517 lGlnLeuArgGlySerTrpArgValGlyGlnThrAlaGlyGlySerArgA 534
|||||
1696 GCAGCTACGGGGTCTTGGAGAGTCGGCCAGCGGGGGGGGCGACGAGGA 1745
|||||
534 snPheAlaSerTyProThrAsnProCysPheProPheSerValProGlu 550
|||||
1746 ACTTGGCTCATACCCACCAACCCCTGCTTCCCTTCTCGTCCCGCAG 1795
|||||
551 GlyProGlyProArgCysValArgIleThrLeuHisGlnHisCysArgPr 567
|||||
1796 GGCCCTGGCCCGCTGCGTCCGCTCATCTCATCATCATCTTCCAGGTCCCG 1845
|||||
567 oSerAspThrGluPheHisProIleGlyPheHisIlePheGlnValProG 584
|||||
1846 CAGTGACACCGAGTTCCACCCATCGCTTCCATATCTTCCAGGTCCCG 1895
|||||
584 lGlyGlyArgSerGlnAspAlaProProLeuLeuGlnGluProLeu 600
|||||
1896 AGGGTGGAGGAGCCAGGAGCCACCCCTGCTGCTGAGGAGCCGCTG 1945
|||||
601 LeuSerCysValProHisArgTrgAlaGlnGluValSerArgLeuCysLe 617
|||||
1946 CTGAGTGGCTGCCACATCGCTACGCCAGGAGGTGAGCGCGCTCTGCT 1995
|||||
617 uLeuProAlaGlyThrTyrlsValValProSerThrTyrlsProAspT 634
|||||
1996 CTGCTGCTGAGGACCTCAGAGTTGTGCCCTCCACCTACTCTCCCGACA 2045
|||||
634 hrGluGlyAlaPheThrValThrIleAlaThrArgIleAspArgProSer 650
|||||
2046 CAGAGGGGGCTTCACAGTGCATGCCAACCAGGATTGACAGGCCATCC 2095
|||||
651 IleHisSerGlnGluMetLeuGlyGlnPheLeuGlnGluValSerValMe 667
|||||
2096 ATTCACAGCCAGGAGATGCTGGGCCAGTTCTCCCAAGAGGTCTCCGTAT 2145
|||||
667 talValMetLysThr 672
|||||
2146 GGCAGTGTGAAAAACC 2161
|||||

seq_name: gb_pr:BC007553

seq_documentation_block:

LOCUS BC007553

DEFINITION Homo sapiens, clone IMAGE:2967328, mRNA, partial cds.

ACCESSION BC007553

VERSION BC007553.1 GI:14043130

KEYWORDS

SOURCE human.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 2614)

AUTHORS

Strausberg, R.

TITLE

Direct Submission

JOURNAL

Submitted (10-MAY-2001) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

REMARK
COMMENT

NIH-MGC Project URL: <http://mgc.mci.nih.gov>
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC/DCTD/DTF
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@cgsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
Ness, Pawan Pandoh, Anna-Jissa Prabhu, Parvaneh Saeedi, Jacqueline
Schein, Duane Smalhus, Michael Smith, Lorraine Spence, Jeff Stott,
Michael Thorne, Miranada Tsai, Natasha van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
Series: IRAL Plate: 22 Row: d Column: 20.

FEATURES

source

1. 2614
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2967328"
/tissue_type="Skin, melanotic melanoma."
/clone_lib="NIH-MGC-20"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
<1..2164
/codon_start=2
/product="Unknown (protein for IMAGE:2967328)"
/protein_id="AA07553.1"
/db_xref="GI:14043131"
/translation="TCGALGAGGAGRNGRGRGPEAAADSPSLRGCDPEATGCRMRP
AEPMRKATPAPARELRFDAAPPAADSSLECDSTPLAOFREDITWRREDEICATPR
LFPDPRREGOVKGLGDCWFLCAGALOKSRHLIDVIPPQSGMADQYRGSFCTP
IWFGRWVEVTTDDRLPLCLAGRLCFRCQREDYVRLLEKVVAKVGSYENHMAQV
ADALVDTGGIAERMLKGVAGSGQDDRPGRMRHRLCGLKDCCLSCCCLSPR
AGARELGEFHAFVSDRLRELQAGAGCILLRIQNPGRRCMOGLMREGGMSQVDA
AVASLLSQOGEFWEFEFEFLFEDELTVGIVTAGHLQSLYTERLCHTRALAG
AMVKGSGAGCCRNNSGFPNPKFWLRVSESEYIVAILOSRLAAMAGARALVSD
SHYSGSPASIPGRHGOAVGLMLKVEKRVNLPRVAMPVAGTACHADYREVLRCE
LSPGYLAVPSTELKDAFGLLKFVETGSAIVAKNTGPALPAGENGTVG
LRSMSRVGTAGSRRNPASYPTNCPPEFVSDEGPRCVRITLHQRDRPDERHPG
FHIQVPEGRSDAPRLLOPELSCVPHRYAOEYSRLCLILPAGTTKVPSTLTPDT
EGAFVTIARIDRPSIHSQEMLQDFLQEVSYMAVMKT"

CDS

BASE COUNT 470 a 809 c 886 g 449 t
ORIGIN

alignment_scores:

Quality: 3649.00 Length: 672
Ratio: 5.430 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-768-877-2 x BC007553 ..

Align seg 1/1 to: BC007553 from: 1 to: 2614

1 MetatgalaGlyArgGlyAlaThrProAlaArgGluLeuPheArgAspAl 17
|||||
146 ATGG 195
17 AlaAlaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrPro 34
|||||
196 CGCCTTCCTCCCGCGGAGCTCTCTCTGCGACTTGTCTACGGCGC 245
34 euAlaGluPheArgGluAspLeuThrThrPArgArgProGlnGluLeuCys 50
|||||
246 TGGCCCATCTCCGGAGGACATCACTGAGGCGGGCCAGAGATTGT 295

51 AlaThrProArgLeuPheProAspAspProArgGluGluValAlaGlySGL 67
|||||
296 GCCACACCCCGGCTGTTCCAGATGACCCAGGGGAGGAGGAGTGAAGCA 345
67 ngILeuLeuGlyAspCysThrPheLeuCysAlaCysAlaAlaLeuGlnL 84
|||||
346 GGGGCTGCTGGGGGATTCGCTGCTCTGCTGCTGCTGCTGCTGCTGCTG 395
84 ySerArgHisLeuLeuAspGlnValIleProProGlnGlnProSerTrp 100
|||||
396 AGACAGAGGACCTCTCGACAGGATATTCCTCGGAGACAGCGAGCTGG 445
101 AlaAspGlnGluThrArgGlySerPheThrCysArgGluLeuPheGln 117
|||||
446 GCCGACAGAGATACCGGGGCTCTTACCTGCGATTTGGCAGCTTTGG 495
117 yArgTrpValGluValThrThrAspAspArgLeuProCysLeuAlaGlyA 134
|||||
496 ACCCTGGGTGAGGTGACACAGATACCGCTGCTGCTGCTGCTGCTGCTG 545
134 rGLeuCysPheSerArgCysGlnArgGluAspValPheThrLeuProLeu 150
|||||
546 GACTCTGTTCTCCCGCTGCGACAGGAGGATGTGTTGCGCTCCCTTA 595
151 LeuGluLysValTyrAlaLysValHisGlySerTyrGluHisLeuTrpAl 167
|||||
596 CTGGAAAGGCTCTACGCCAAGGTCATGGGTCTCTACGACACCTGTGGGC 645
167 acGlyGlnValAlaAspAlaLeuValAspLeuThrGlnGlyLeuAlaGluA 184
|||||
646 CCGGCGAGGTGGGGGATGCCCTGGTGAACCTGACCGGGGCTGGCGAAGA 695
184 rGTTPAsnLeuLysGlyAlaAlaGlySerGlyGlyGlnGlnAspArgPro 200
|||||
696 GATGAAACCTGAAGGCGGTACAGCAGAAAGCGAGGACAGGACAGGCA 745
201 GlyArgTrpGlnHisArgThrCysArgGlnLeuLeuHisLeuLysAspGL 217
|||||
746 GGGCGGTGGAGACACAGGACTTGTGCGAGCTGCTCACCTTAAGGACCA 795
217 nCysLeuIleSerCysCysValLeuSerProArgAlaGlyAlaArgGluL 234
|||||
796 GGTGTGATCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 845
234 euGlyGluPheHisAlaPheIleValSerAspLeuArgGluLeuGlnGly 250
|||||
846 TGGGGGAGTTCCATGCTCATTTGCTCGAGCTGGGAGGAGTCCAGGCT 895
251 GlnAlaGlyGlnCysIleLeuLeuLeuArgGlnGlnAspProTrpGlyAr 267
|||||
896 CAGCGGGCCAGTGCATCTCTGCTGCGATCCAGAACCCCTGGGGCGG 945
267 gATGcysTrpGlnGlyLeuTrpArgGluGlyGlyGluGlyTrpSerGln 284
|||||
946 GCGGTGCTGGGAGGGGCTTGGAGAGAGGGGGGTGAAGGCTGGAGCCAGG 995
284 AlaAspAlaAlaValAlaSerGluLeuLeuSerGlnLeuGlnGluGlu 300
|||||
996 TAGATGACAGCGGTACATCTGAGCTCTGCTGCTGCTGCTGCTGCTGCTG 1045
301 PheTrpValGluGluGluGluPheLeuArgGluLeuAspGlnLeuThrVa 317
|||||
1046 TTCTGGGTGAGAGAGAGAGATTCTCTCAGGAGTTTGAAGACCTCAACCT 1095
317 IGlyTyrProValThrGluAlaGlyHisLeuGlnSerLeuTyrThrGluA 334
|||||
1096 TGGTACCCCGGTACAGGAGGCGGCGCACCTGACAGACCTTACACAGAGA 1145
334 rGLeuLeuCysHisThrArgAlaLeuProGlyAlaTrpValLysGlyGln 350
|||||
1146 GGCTGCTCTGCGCATACGCGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1195

351 SerAlaGlyGlyCysArgAsnAsnSerGlyPheProSerAsnProLysPhe 367
|||||
1196 TCAGCAGAGGCTGCGGGAACAACAGCGGCTTCCAGCAACCCCAATT 1245
367 eTrpLeuArgValSerGluProSerGluValTyrIleAlaValLeuGlnA 384
|||||
1246 CTGCTGCGGGTCTCAGAACCGAGTGGGTGATCATTTGCGTCTCGAGA 1295
384 rGSerArgLeuHisAlaAlaAspTrpAlaGlyArgAlaArgAlaLeuVal 400
|||||
1296 GATCCAGGCTGCACGCGGCGACTGGGCGCGCGGCGGCGGCACTGGTG 1345
401 GlyAspSerHisThrSerTrpSerProAlaSerIleProGlyLysHisTy 417
|||||
1346 GGTGACAGTCATCTTCTGAGAGCCAGCGAGCATCCCGGCAAGCACTA 1395
417 rGlnAlaValGlyLeuHisLeuTrpLysValGluLysArgArgValAsnL 434
|||||
1396 CGAGGCTGGGTCTGCACCTCTGGAGGTAGAGAGCGGCGGTCAATC 1445
434 euProArgValLeuSerMetProProValAlaGlyThrAlaCysHisAla 450
|||||
1446 TGCTAGGTCCTGTCATGCCCCCGTGGCTGGCACCCTGTCATGCA 1495
451 TyrAspArgGluValHisLeuArgCysGluLeuSerProGlyTyrTyrLe 467
|||||
1496 TAGCAGCGGAGGTCACCTGCTGCTGTGAGCTCTACCGGCTACTACCT 1545
467 uAlaValProSerThrPheLeuLysAspAlaProGlyGluPheLeuA 484
|||||
1546 GGCTGTCCCGCAGCAGCTTCTGAGAGCGCGCGGGGAGTTCCTGCTCC 1595
484 rGValPheSerThrGlyArgValSerLeuSerAlaIleArgAlaValAla 500
|||||
1596 GAGTCTTCTCTACCGGGCGAGTCTCCCTTAGCGCCATCAGGCGAGTGCC 1645
501 LysAsnThrThrProGlyAlaAlaLeuProAlaGlyGluTrpGlyThrVa 517
|||||
1646 AGAACACCAACCCCGGGGAGCCCTGCTGCGGGGAGTGGGGACCGT 1695
517 lGlnLeuArgGlySerTrpArgValGlyGlnThrAlaGlyGlySerArgA 534
|||||
1696 GCAGCTACGGGGTCTTGAGAGAGTCGCCAGACGGCGGGGGGAGCAGGA 1745
534 snPheAlaSerTyrProThrAsnProCysPheProPheSerValProGlu 550
|||||
1746 ACTTTGCTCATACCCACCAACCCCTGCTTCCCTTCTCGGTCCCGAG 1795
551 GlyProGlyProArgCysValArgIleThrLeuHisGlnHisCysArgPr 567
|||||
1796 GGCCCTGGCCCCCGTGGCTCCGCATCACTCTGCATCAGCATGCGCGCC 1845
567 oSerAspThrGluPheHisProIleGlyPheHisIlePheGlnValProG 584
|||||
1846 CAGTGACACCGAGTTCACCCCTCGGCTTCCATATCTTCCAGGTCCAG 1895
584 luGlyGlyArgSerGlnAspAlaProProLeuLeuLeuGlnGluProLeu 600
|||||
1896 AGGGTGAAGGAGCAGGAGCAGCACCACCTGCTGCTGCGAGGCGCGTG 1945
601 LeuSerCysValProHisArgTyrAlaGlnGluValSerArgLeuCysLe 617
|||||
1946 CTGAGCTGGTCCACATCGCTACGCCCGGAGGTGAGCGGCTCTGCT 1995
617 uLeuProAlaGlyThrTyrLysValValProSerThrTyrTrpLeuProAspT 634
|||||
1996 CTGCTGCAGGACCTACAGAGTTGTGCTTCCCTCCACCTACCTGCGGACA 2045
634 hrGluGlyAlaPheThrValThrIleAlaThrArgIleAspArgProSer 650
|||||
2046 CAGAGGGGGCTTTCAGTGCACCATCGCAACCCAGGATTGACAGGCCATCC 2095
651 IleHisSerGlnGluMetLeuGlyGlnPheLeuGlnValSerValIle 667

|||||
2096 ATTCAGAGCCAGAGATGCTGGCCAGTTCTCCAGAGAGGTCTCCGTCAT 2145
667 talValMetLysThr 672
|||||
2146 GGCAGTGATGAAACC 2161
seq_name: gb_pat:AR153290
seq_documentation_block:
LOCUS AR153290 2620 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 3 from patent US 6235481.
ACCESSION AR153290
VERSION AR153290.1 GI:15120822
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2620)
AUTHORS Horikawa,Y., Oda,N., Hanis,C.L., Bell,G.I. and Cox,N.J.
TITLE Polynucleotides encoding calpain 10
JOURNAL Patent: US 6235481-A 3 22-MAY-2001;
FEATURES
source
1. 2620
/organism="unknown"
BASE COUNT 448 a 819 c 899 g 454 t
ORIGIN
alignment_scores:
Quality: 3649.00 Length: 672
Ratio: 5.430 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-768-877-2 x AR153290 ..
Align seg 1/1 to: AR153290 from: 1 to: 2620
1 MetArgAlaGlyArgGlyAlaThrProAlaArgGluLeuPheArgAspAl 17
178 ATGCGGGCGGGCGGGCGGCGAGCGCGGAGGAGCTGTTCCGGGACGC 227
17 aAlaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrProL 34
228 CGCTTCCCGCGCGGACTCTCGCTCTTCTGCGACTTGTCTACGCGCG 277
34 euAlaGlnPheArgGluAspIleThrTrpArgArgProGlnGluIleCys 50
278 TGGCCCAAGTTCCGCGAGGACATCAGTGGAGGCGGCCCGGAGATTGT 327
51 AlaThrProArgLeuPheProAspAspProArgGluGlyGlnValLysGl 67
328 GCCACACCCCGGCTGTTTCCAGATGACCCAGGGAAGGCGAGTGAAGCA 377
67 nGlyLeuLeuGlyAspCysTyrPheLeuCysAlaCysAlaAlaLeuGlnL 84
378 GGGGCTGCTGGGGGATTGCTGTTCTCTGCTGCTGCTGCTGCTGCTG 427
84 ysSerArgHisLeuLeuAspGlnValIleProProGlyGlnProSerTrp 100
428 AGAGCAGCACCTCTCTGGCACCAGGTTCATCTCCGGGACAGCCGAGCTGG 477
101 AlaAspGlnGluTyrArgGlySerPheThrCysArgIleTrpGlnPheGl 117
478 GCGGACAGGAGTACCGGGGCTCTCTTCCACCTCTGCGCATTTGGCAGTTGG 527
117 yArgTrpValGluValThrThrAspArgLeuProCysLeuAlaGlyA 134
528 ACGTGGGTGGAGGTGACACAGATGACCGCTGCGGTGCTGCGGGA 577
134 rGLeuCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
|||||

578 GACTCTGTTTCCCGCTCCAGAGGAGATGTTCTGGCTCCCTTA 627
 151 LeuGluValValTyrAlaValHisGlySerTyrGluHisLeuTyrAla 167
 628 CTGGAAGAGTCTTACGCCAGGTCATGGTCTTACAGACACTGTGGGC 677
 167 aglyValValAlaAspAlaValAspLeuThrGlyGlyLeuAlaGlu 184
 678 CGGGAGGTGGCGGATGGCTGGTGGACTGACCGCGGCTGGCAGAAA 727
 184 rgtPAspLeuGlyValAlaGlySerGlyGlyGlnAspArgPro 200
 728 GATGGAACTTGAAGGCTTACAGAACGGAGCCACAGACAGGCCA 777
 201 GlyArgTyrPgluHisArgThrCysArgGlnLeuHisLeuLysAspG 217
 778 GGGCGCTGGACACAGACTTGTGGCAGCTGCTCCACTGAAGAGCCA 827
 217 nCysLeuLeuSerCysValLeuSerProArgAlaGlyAlaArgGln 234
 828 GTGTCTGATCAGCTGCTGCTGCTCAGCCACAGACAGTGGCGGAGC 877
 234 euGlyGluPheHisAlaPheIleValSerAspLeuArgGlnLeuGln 250
 878 TGGGGAGATTCCATGCCCTTCAATGTCTCGACCTGGAGCTCCAGGGT 927
 251 GlnAlaGlyGlnCysIleLeuLeuLeuArgIleGlnAspProTyrPgl 267
 928 CAGCGGGGCCAGTGCATCTCTGCTGGATCCAGAACCCCTGGGGCG 977
 267 gArgCysTyrPglGlnGlyLeuTyrArgGlyGlyGlyLeuTyrPserGln 284
 978 GCGGTCTGGAGGGGCTCTGAGAGAGGGGGTGAAGGTGAGGAGCCAG 1027
 284 alaAspAlaAlaValAlaSerGluLeuLeuSerGlnLeuGlnGlyGlu 300
 1028 TAGATGACAGCGGTACATGTGCTCTCCAGCTCCAGGAAGGGGAG 1077
 301 PheTyrValGluGluGluPheLeuArgGlnPheAspGluLeuThrVa 317
 1078 TTCTGGGGGAGAGAGAGACTTCTCAGAGTTTACAGACTCACCGT 1127
 317 IGIYTrProValThrGluAlaGlnHisLeuGlnSerLeuTyrThrGlu 334
 1128 TGGCTACCCGGTACAGGAGCGCGCCACCTCAGAGCTTACACAGGA 1177
 334 rGluLeuLysHisThrArgAlaLeuProGlyAlaTyrValLysGlyGln 350
 1178 GCGTGTCTTGGCATACGGGGCGCTGGCGCTGGGTCAAGGGCCAG 1227
 351 SerAlaGlyGlyCysArgAsnAsnSerGlyPheProSerAsnProLysP 367
 1228 TCAGCAGAGGCTCCCGAACACACAGCGCTTCTCCAGACACCCCAAT 1277
 367 eTyrLeuArgValSerGluProSerGluValTyrIleAlaValLeuGln 384
 1278 CTGGCTCGGGGTCTCAGAACCGACTGAGGTATGATGGCTCTTGAGA 1327
 384 rGSerArgLeuHisAlaAlaAspTyrAlaGlyArgAlaValLeuVal 400
 1328 GATCCAGGCTGCACGCGCGGCTGGCGAGCCCGGGCCGCTGGTG 1377
 401 GlysPserHisThrSerTyrProAlaSerIleProGlyLysHisTyr 417
 1378 GGTACAGTACTACTTGTGTGAGCCCAAGACATCCCGGCGCAAGCACTA 1427
 417 rGlnAlaValGlyLeuHisLeuTyrLysValGluLysArgValAsnL 434
 1428 CCAAGGCTGTGGTCTGCACCTCTGAAGGTGAGAACGGGGGGTCAATC 1477
 434 euProArgValLeuSerMetProProValAlaGlyThrAlaCysHisAla 450
 1478 TGCCTAGGGTCTGTCCATGCCCGCTGGCTGGCACCGGGTGGCCATGCA 1527

451 TyrAspArgGluValHisLeuArgCysGlnLeuSerProGlyTyrTyrLe 467
 1528 TACAGCCGGAGGCTCCACCTTCTGTGAGCTTCCACCGGGCTACTACT 1577
 467 uAlaValProSerThrPheLeuLysAspAlaProGlyGluPheLeu 484
 1578 GGCTGTCCCAAGCACTTCTGAAGAGCGCCAGGGAGTTCTGTCTCC 1627
 484 rValPheSerThrGlyArgValSerLeuSerAlaIleArgAlaValAla 500
 1628 GAGCTTCTTACCGGGGAGTCTCCCTTACGGCATAGGCGAGTGGCC 1677
 501 LysAsnThrThrProGlyAlaAlaLeuProAlaGlyTyrPglTyrVa 517
 1678 AAGAACACACCCCGGGGAGCCCTGCTCGGGGGAGTGGGAGCCGT 1727
 517 IGIleuArgGlySerTyrPglValGlyGlnThrAlaGlyLysSerArg 534
 1728 GCACCTACGGGCTTCTTGAGAGTCCGCCAGACGGGGGGGAGACAGA 1777
 534 snPheAlaSerTyrProThrAsnProCysPheProPheSerValProGlu 550
 1778 ACTTGCCTCATACCCCAACACCCCTGCTTCCCTTCTGCTCCGAG 1827
 551 GlyProGlyProArgCysValArgIleThrLeuHisGlnHisCysArgPr 567
 1828 GGGCTGGCCCCCGCTGCTCGCATCATCTGCATCAGCACTCCGGCC 1877
 567 oSerAspThrGluPheHisProIleGlyPheHisIlePheGlnValProG 584
 1878 CAGTGACACGAGTCCACCCCATCGGCTTCATATCTTCAGAGTCCAG 1927
 584 IuGlyGlyArgSerGlnAspAlaProProLeuLeuLeuGlnGluProLeu 600
 1928 AGGCTGAGAGGACGACAGACGACCCCACTGCTGTCAGAGACCGCTG 1977
 601 LeuSerCysValProHisArgTyrAlaGlnGluValSerArgLeuCysLe 617
 1978 CTGAGCTGCGTCCCATCGCTACGCCAGAGGTGAGCGGCTCTGCT 2027
 617 uLeuProAlaGlyThrTyrLysValValProSerThrTyrLeuProAsp 634
 2028 CCTCGTCGACGACCTACAAAGTTTGGCTTCCACTACTGCGGAGCA 2077
 634 hrGluGlyAlaPheThrValThrIleAlaThrArgIleAspArgProSer 650
 2078 CAGAGGGGGCTTACAGTGCATGCCAACCAAGATGAGACAGCCATCC 2127
 651 IleHisSerGlnIleMetLeuGlyGlnPheLeuGlnGluValSerValMe 667
 2128 ATTCAACAGCAGAGAGTGTGGGCACTTCTCCAGAGGTCTCGTCAT 2177
 667 tAlaValMetLysThr 672
 2178 GGCAGTATGATAAAC 2193
 seq_name: gb_pr:AF089088
 seq_documentation_block: 2620 bp mRNA linear pri 23-JAN-2001
 LOCUS AF089088
 DEFINITION Homo sapiens calpain-like protease CAPN10a mRNA, complete cds.
 ACCESSION AF089088
 VERSION AF089088.1 GI:10503934
 KEYWORDS
 SOURCE
 ORGANISM
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 2620)
 Horikawa,Y., Oda,N., Cox,N.J., Li,X., Orho-Melander,M., Hara,M.,
 Hinokio,Y., Lindner,T.H., Mashima,H., Schwarz,P.E., del
 Bosque-Plata,L., Horikawa,Y., Oda,Y., Yoshuchi,I., Collina,S.,

Polonsky, K.S., Wei, S., Concannon, P., Iwasaki, N., Schulze, J.,
Baier, L.J., Bogardus, C., Groop, L., Boerwinkle, E., Hanis, C.L. and
Bell, G.I.
Genetic variation in the gene encoding calpain-10 is associated
with type 2 diabetes mellitus
Nat. Genet. 26 (2), 163-175 (2000)
20472315
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
Submitted (02-SEP-1998) Howard Hughes Medical Institute, The
University of Chicago, 5841 S. Maryland Avenue, MC1028, Chicago, IL
60637, USA

FEATURES
source
1..2620
/organism="Homo sapiens"
/db.xref="taxon:9606"
178..2196
/note="type A"
/codon_start=1
/product="calpain-like protease CAPN10a"
/protein_id="AAG17966.1"
/db.xref="GI:10503935"
/translation="MRAGGATPARELFRDAFPADSSLFCDLSTPLAQREDITWR
RPEICATPRFDDPREGQVKGLGDCWFLCACAALQKSRHLLDQVIPGPQSWAD
QYRGSTFCRIWQGRWVEVTDRLPLAGRLCFSCOREDFWLPLEKRYAKVHG
SYELHAGQVADALVDLTGGLAERNLKGVAGSGQDQDPRGRWEHRTCOLLLKDDC
LISCCVSPRAGARELGEFHAFIVSDRLRELQAGQCTILLRIQNPGRRCWQGLWRE
GEGMSQDAAVASSELLSQEGEFVEEFELREFDELTVGVPTVTEAGHLQSLYTER
LLCHTRALPCAWKVGSGAGGRNNSGPPNPKFWLRVSEFVYIAVLQSRLLHAADW
AGRALVGDSTSWSPASIPGKHQVQLHWKVRKRNVLPRVLSMPVPVAGTACHA
YDREHLRCELSPGYLAVPSTFLKDPGEFLRVSTGRVSLSAIRAVAKNTTPGAA
LPAGEWTVQLRGWSRGVQTAGGSRNFASTPNCFPSVPEGPGRCVRITLHQHCR
PSDEFPIGFHFQVPEGGSDAPLLQEPLELLSCVPHRYAOEVSRLCLLPAGTVK
VVPSTYLPDTEGAFVTIATRIDRPSIHSQEMQLQFLEQVSMVAMKT"
polyA_signal
BASE COUNT 448 a 819 c 899 g 454 t
ORIGIN

alignment_scores
Quality: 3649, 00 Length: 672
Ratio: 5,430 Gaps: 0
Percent Similarity: 100,000 Percent Identity: 100,000
alignment_block
US-09-768-877-2 x AF089088 ..
Align seg 1/1 to: AF089088 from: 1 to: 2620
1 MetArgAlaGlyArgGlyAlaThrProAlaArgGluLeuPheArgAspAl 17
178 ATGCGGGCGGGCGGGCGGCGAGCGCGGAGGAGCTGTCGGGAGCGC 227
17 aAlaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrProL 34
228 CGGCTTCCCGCGCGGAGCTCTCGCTCTTCGCGACTTGTCTACGCGCG 277
34 euAlaGlnPheArgGluAspIleThrTrpArgProGlnGluIleCys 50
278 TGCGCCAGTTCGCGAGGACATCAGTCGAGGCGGCGGCCCGGAGATTGT 327
51 AlaThrProArgLeuPheProAspAspProArgGluGlnValIysG 67
328 GGCACACCCCGCGGTGTTCCAGATGACCCAGCGGAGGCGAGTGAAGCA 377
67 nGlyLeuLeuGlyAspCysTrpPheLeuCysAlaCysAlaAlaLeuGlnL 84
378 GGGGCTGCTGGGGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 427
84 ySerArgHisLeuLeuAspGlnValIleProProGlyGlnProSerTrp 100
428 AGAGCAGGCACCTCTCTGGACACAGTCAATTCCTCCGGGACAGCGGAGTGG 477

101 AlaAspGlnGluTrpArgGlySerPheThrCysArgIleTrpGlnPheG 117
478 GCCGACGAGAGTACCGGGCTCTTACCTGTCGCAATTGGCAGTTGG 527
117 YArgTrpValGluValThrThrAspAspArgLeuProCysLeuAlaGly 134
528 ACCTGGGTGGAGGTGACACAGATGACCGCTGCGCTGCTTGCAGGGA 577
134 rGLeuCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
578 GACTCTGTTCTCCCGCTGCCAGAGGAGGATGTTCTGGCTCCCTTA 627
151 LeuGluLysValTrpAlaLysValHisGlySerTyrGluHisLeuTrpAl 167
628 CTGGAAGAAGGTCTACGCCAAGTCCATGGTCTCTACAGCACCTGTGGGC 677
167 aGlyGlnValAlaAspAlaLeuValAspLeuThrGlyGlyLeuAlaGlu 184
678 CGGGCAGGTGGCGGATGCTGCTGTCGACCTGACCGCGCGCTGGCAGAA 727
184 rGTrpAsnLeuLysGlyValAlaGlySerGlyGlnGlnAspArgPro 200
728 GATGGAACCTGAAGGCGGTAGCAGGAAGCGGAGCGGAGCAGCAGGCA 777
201 GlyArgTrpGluHisArgThrCysArgGlnLeuLeuHisLeuLysAspG 217
778 GCGCGCTGGGACACAGGACTTGTGCGCAGCTCTCCACTGAAGGACA 827
217 nCysLeuIleSerCysCysValLeuSerProArgAlaGlyAlaArgGluL 234
828 GTGCTCTGATCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 877
234 euGlyClnPheHisAlaPheIleValSerAspLeuArgGluLeuGlnGly 250
878 TGGGGGAGTTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 927
251 GlnAlaGlyGlnCysIleLeuLeuLeuArgIleGlnAsnProTrpGlyAr 267
928 CAGCGGGCGGAGTGCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 977
267 gArgCysTrpGlnGlyLeuTrpArgGluGlyGlyGlyGlyTrpSerGln 284
978 GCGGTGCTGGCAGGGGCTCTGAGAGAGGGGGGTGAAGGCTGAGCCAG 1027
284 aAlaAspAlaValAlaSerGluLeuLeuSerGlnLeuGlnGluGlyGlu 300
1028 TAGATGACCGGTAGCATCTGAGCTCTCTGCTGCTGCTGCTGCTGCTG 1077
301 PheTrpValGluGluGluPheLeuArgGluPheAspGluLeuThrVa 317
1078 TTTCTGGGTGGAGGAGGAGGAGTTCCTCAGGAGTTTGACGAGCTCACCT 1127
317 lGlyTrpProValThrGluAlaGlyHisLeuGlnSerLeuTrpThrGlu 334
1128 TGCTTACCGGTCACGGAGCGCGCGGCTGCTGCTGCTGCTGCTGCTGCT 1177
334 rGLeuLeuCysHisThrArgAlaLeuProGlyValaTrpValLysGlyGln 350
1178 GGCTGCTCTGCCATAGCGGGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 1227
351 SerAlaGlyGlyCysArgAsnAsnSerGlyPheProSerAsnProLysPh 367
1228 TCAGCAGGAGGCTGCCGGAACAACAGCGGCTTTTCCAGCAACCCCAATT 1277
367 eTrpLeuArgValSerGluProSerGluValTrpIleAlaValLeuGln 384
1278 CTGGCTGCGGGTCTCAGAACCGAGTGAGTGATGCTGCTGCTGCTGCTG 1327
384 rGSerArgHisAlaAlaAspTrpAlaGlyArgAlaArgAlaLeuVal 400
1328 GATCCAGGCTGACCGCGGGGACTGGGAGGCGCGGGGCGGCGGCGGCT 1377

```

401 GlyAspSerHisThrSerTrpSerProAlaSerIleProGlyLysHisTyr 417
|||||
1378 GGTGACAGTACTACTCGTGGAGCCAGGACCATCCGGGAGACACACTA 1427
|||||
417 TCTAAlaValAlaGlyLeuHisLeuTrpLysValGluLysArgArgValAsnL 434
|||||
1428 CCAGGCTGTGGGTGTGCACCTCTGAGAAAGGTAGAAAGCGCGGCTCAATC 1477
|||||
434 euProArgValLeuSerMetProProValAlaGlyThrAlaCysHisAla 450
|||||
1478 TGCCCTAGGGTCTCTGTCATGCCCCCGGTGGCTGGACCGCGGTGCATGCA 1527
|||||
451 TyrAspArgGluValHisLeuArgCysGluLeuSerProGlyTyrTyrLe 467
|||||
1528 TAGACCGGGAGGTCCACCTCGTTGTGAGCTCTCACCGGCTACTACTCT 1577
|||||
467 uAlaValProSerThrPheLeuLysAspAlaProGlyGluPheLeuLeu 484
|||||
1578 GGCTGTCCCGACACCTTCTCTGAAAGAGCGCCAGGGGAGTTCCTGCTCC 1627
|||||
484 rGValPheSerThrGlyArgValSerLeuSerAlaIleArgAlaValAla 500
|||||
1628 GAGTCTCTCTACGGGGGAGGTCTCCCTTACGCGCATCAGGCGAGTGGCC 1677
|||||
501 LysAsnThrThrProGlyAlaAlaLeuProAlaGlyGluTrpGlyThrVa 517
|||||
1678 AAGAACACACACCCCGGGGAGACCCCTGCTGGGGGAGTGGGGAGACCT 1727
|||||
517 IGlLeuArgGlySerTrpArgValGlyGlnThrAlaGlyLysArgAr 534
|||||
1728 GCAGCTAGGGGTTCTGTGAGAGTGGCCAGACGGGGGGGAGACAGAGA 1777
|||||
534 snPheAlaSerTyrProThrAsnProCysPheProPheSerValProGlu 550
|||||
1778 ACTTGGCTCATACCCACCAACCCCTGCTTCCCTTCTCGGTCCCGAG 1827
|||||
551 GLyProGlyProArgCysValArgIleThrLeuHisGlnHisCysArgPr 567
|||||
1828 GGCCCTGGCCCCCGCTGGCTCCGATCCTGATCATGACACTGGCGGCC 1877
|||||
567 oSerAspThrGluPheHisProIleGlyPheHisIlePheGlnValProG 584
|||||
1878 CAGTGACACCGAGTTCACACCCCATGCTTCATATCTTCCAGGCTCCAG 1927
|||||
584 IuGlyGlyArgSerGlnAspAlaProProLeuLeuGlnGluProLeu 600
|||||
1928 AGGGTGGAGAGAGACCAAGACCAACCCCATGCTGCTCAGAGAGCGCTG 1977
|||||
601 LeuSerCysValProHisArgTyrArgIleGlnGluValSerArgLeuCysLe 617
|||||
1978 CTGAGCTCGTGGCCATCGCTAGCCGAGGAGGTGAGCCGGCTGTGCT 2027
|||||
617 uLeuProAlaGlyThrTyrLysValValProSerThrTyrLeuProAspT 634
|||||
2028 CCTGCTGTGCAGGACCTACAGAGTGTGCTCTCCACTACCTGCGCGACA 2077
|||||
634 hrGluGlyAlaPheThrValThrIleAlaThrArgIleAspArgProSer 650
|||||
2078 CAGAGGGGGCTTCTCAGCTGACCATGCAACCCAGGATTGACAGGCAACC 2127
|||||
651 IleHisSerGlnGluMetLeuGlyGlnPheLeuGlnGluValSerValMe 667
|||||
2128 ATTCAACAGCCAGAGATGCTGGGCCAGTTCCTCCAAAGAGTCTCCGTAT 2177
|||||
667 talValMetLysThr 672
|||||
2178 GGCAGTCAATGAAGAAC 2193
|||||
seq_name: gb_pat:ARI53295
seq_documentation_block: 2455 bp DNA linear PAT 08-AUG-2001
LOCUS ARI53295
DEFINITION Sequence 13 from patent US 6235481.

```

```

ACCESSION ARI53295
VERSION ARI53295.1 GI:15120827
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2455)
AUTHORS Horikawa, Y., Oda, N., Hanis, C. L., Bell, G. I. and Cox, N. J.
TITLE Polynucleotides encoding calpain 10
JOURNAL Patent: US 6235481-A 13 22-MAY-2001;
FEATURES
source 1..2455
BASE COUNT 429 a 769 c 827 g 430 t
ORIGIN

alignment_scores:
Quality: 3604.00 Length: 673
Ratio: 5.395 Gaps: 2
Percent Similarity: 99.257 Percent Identity: 99.257

alignment_block:
US-09-768-877-2 x ARI53295
Align seg 1/1 to: ARI53295 from: 1 to: 2455

1 MetArgAlaGlyArgGlyAlaThrProAlaArgGluLeuPheArgAspAl 17
|||||
42 ATGCGCGCGGGCGCGGGCGCGAGCGCGCGGAGGAGAGTGTTCGGGAGCC 91
|||||
17 aAlaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrProL 34
|||||
92 CGCCTTCCCGCGCGGAGCTCTGCTCTTGTGCACTTGTGTACGCCGC 141
|||||
34 euAlaGlnPheArgGluAspIleThrTrpArgArgProGlnGluLys 50
|||||
142 TGCCCGAGTTCGCGAGGACATCAGTGGAGCGGCCCGCCAGAGATTCT 191
|||||
51 AlaThrProAlaGluPheProAspAspProArgGluGlyGlnValLysG 67
|||||
192 GCCACACCCCGCGCTTTCAGATGACCAACCGAGGAGGAGTGAAGCA 241
|||||
67 nGlyLeuGluGlyAspCysTrpPheLeuCysAlaCysAlaAlaLeuGln 84
|||||
242 GGGGCTGTGGGGGATTCGTGCTTCTGTGCTCTGCGCGCGGCTGCAGA 291
|||||
84 ySerArgHisLeuLeuAspGlnValIleProProGlyGlnProSerTrp 100
|||||
292 AGAGAGAGCACTCTCTGAGACCAAGTCAATTCCTCCGGACAGCCGAGCTG 341
|||||
101 AlaAspGlnGluTyrArgIleSerPheThrCysArgIleTrpGlnPheG 117
|||||
342 GCCGACCAAGAGTACCGGGGCTCTTCACTGCTGCATTGGCATTTGG 391
|||||
117 YArgTrpValGluValIleThrAspAspArgLeuProCysLeuAlaGly 134
|||||
392 ACGCTGGGTGGAGGTGACACAGATGACCGCTGCGCTTGCCTTGACGGA 441
|||||
134 rGLeuCysPheSerArgCysGlnArgGluAspValPheThrLeuProLeu 150
|||||
442 GACTGTGTTTCTCCGCTGCCAGAGGAGAGATGTTCCTGCTCCCTTA 491
|||||
151 LeuGluLysValTyrAlaLysValHisGlySerTyrGluHisLeuTrp 167
|||||
492 CTGGA.....AAGGTTCATGGGTCTTACAGACACCTGTGG 529
|||||
167 IaglyGlnValAlaAspAlaLeuValAspLeuThrGlyGlyLeuAlaGlu 183
|||||
530 CCGGCGAGGTGGCGGATGCCCTGTGAGACTACCGCGCGGCTGCGAGAA 579
|||||
184 ArgTrpAsnLeuLysGlyValAlaGlySerGlyGlnGlnAspArgPr 200
|||||

```

580 AGATGGAACTGAAGGGCGTAGCAGGAAGCGGAGCCAGCAGGACAGGCC 629
200 OGIYArgTrpGluHisArgThrCysArgGlnLeuLeuHisLeuLysAspG 217
630 AGGCCGCTGGAGACACAGACTGTGCGCAGCTGCTCCACCTGAGGACC 679
217 lncysLeuileSerCysCysValLeuSerProArgAlaGlyAlaArgGlu 233
680 AGTCTCTGATCAGCTGCTGCGTCTCAGCCAGCCAGAGCAGCTGCCGGAG 729
234 LeuGlyGluPheHisAlaPheIleValSerAspLeuArgGluLeuGlnI 250
730 CTGGGGAGTTCATGCTTCATGTCTCGGACCTCGGGAGCTCCAGGG 779
250 YGlnAlaGlyGlnCysIleLeuLeuArgIleGlnAsnProTrpGlyA 267
780 TCAGCGGGCCAGTCATCTGCTGCTGCGATCCAGAACCCCTGGGGCC 829
267 rArgCysTrpGlnGlyLeuTrpArgGluGlyGlyGlyTrpSerGln 283
830 GCGCGTCTGCGCAGGGGCTCTGAGAGAGAGGGGCTGAAGGTTGAGCCAG 879
284 ValAspAlaAlaValAlaSerGluLeuLeuSerGlnLeuGlnGlyG 300
880 GTAGATGCAGCGTAGCATCTGAGCTCTGCTCCAGCTCCAGGAAGGGA 929
300 upheTrpValGluGluGluPheLeuArgGluPheAspGluLeuThry 317
930 GTTCTGGTGGAGAGGAGGAGTCTCAGGAGTTTGAGCAGCTCACCG 979
317 aGlyTrpProValThrGluAlaGlyHisLeuGlnSerLeuTrpThry 333
980 TTGGCTACCGGTCACGGAGGCGCGCACCTGCAGAGCGCTCTACACAG 1029
334 ArgLeuLeuCysHisThrArgAlaLeuProGlyAlaTrpValLysGly 350
1030 AGGCTGCTCTGCCATAGCGGGCGCTGCTTGGGGCTTGGTCAAGGGCCA 1079
350 nSerAlaGlyGlyCysArgAsnAsnSerGlyPheProSerAsnProLys 367
1080 GTCAGCAGAGGCTGCGGAAACAACAGCGCTTTCAGCAACCCCAAT 1129
367 heTrpLeuArgValSerGluProSerGluValTrpIleAlaValLeuG 383
1130 TCTGGCTGGGGTCTCAGAACCGAGTGAGTGTCATTTGCCGCTCCAG 1179
384 ArgSerArgLeuHisAlaAlaAspTPAlaGlyArgAlaAlaLeuVa 400
1180 AGATCCAGGCTGCACCGCGGACTGGCAGGCGCGCGCGGCTGGT 1229
400 lGlyAspSerHisThrSerTrpSerProAlaSerIleProGlyLysHis 417
1230 GGTGACAGTCATCTCTGTCGAGCCAGCAGCATCCCGGCGAGCACT 1279
417 YTGlnAlaValGlyLeuHisLeuTrpLysValGluLysArgValAsn 433
1280 ACCAGCTGTGGCTGTCACCTCTGGAAGGTAGAGAGCGGGGTCAT 1329
434 LeuProArgValLeuSerMetProProValAlaGlyThrAlaCysHisAl 450
1330 CTGCTAGGCTCTGTCCATGCCCCCGTGGTGGCAGCGCGTGCATGC 1379
450 aTyrAspArgGluValHisLeuArgCysGluLeuSerProGlyTrpTyrL 467
1380 ATACGACCGGGAGTCCACTCGTGTGAGCTCTCAGCGGCTACTACC 1429
467 euAlaValProSerThrPheLeuLysAspAlaProGlyGluPheLeu 483
1430 TGGCTGCTCCCGACAGCTTCTTGAAGGACGCGCGAGGAGTCTCTGCTC 1479
484 ArgValPheSerThrGlyArgValSerLeuSerAlaIleArgAlaValAl 500
1480 CGAGTCTTCTACCGGGCGAGTCTCCCTTAGCGCCATCAGGCGAGTGGC 1529

500 aLysAsnThrThrProGlyAlaAlaLeuProAlaGlyGluTrpGlyThrv 517
1530 CAAGACACACCCCGGGGACGCCCTGCTGCGGGGAGTGGGGAGCC 1579
517 alGlnLeuArgGlySerTrpArgValGlyGlnThrAlaGlyGlySerArg 533
1580 TGCAGCTACGGGTTCTTGGAGAGTGGCGAGCAGCGGGGGGCGAGG 1629
534 AsnPheAlaSerTyrProThrAsnProCysPheProPheSerValProG 550
1630 AACTTTGCTCATACCCCAACCCCTGCTCCCTTCTCGTCCCGCA 1679
550 uGlyProGlyProArgCysValArgIleThrLeuHisGlnHisCysArgP 567
1680 GGGCCCTGGCCCCCGCTGCTGCGCATCACTGTCATCAGCACTGCGGC 1729
567 roSerAspThrGluPheHisProIleGlyPheHisIlePheGlnValPro 583
1730 CCAGTGACACCGAGTTCACCCCATCGCTTCCATATCTTCCAGGTCCA 1779
584 GluGlyGlyArgSerGlnAspAlaProProLeuLeuGlnGluProLe 600
1780 GAGGGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1829
600 uLeuSerCysValProHisArgTyrAlaGlnGluValSerArgLeuCysL 617
1830 GCTGAGCTGCTGCGCATCGCTTACGCCAGGAGGTGAGCGGCTCTGCC 1879
617 euLeuProAlaGlyThrTyrLysValValProSerThrTyrLeuProAsp 633
1880 TCTGCTGTCAGGACCTACAGGTTGTGCTTCCCTTCCACTTACCTGCCG 1929
634 ThrGluGlyAlaPheThrValThrIleAlaThrArgIleAspArgProse 650
1930 ACAGAGGGGGCTTCCAGCTGACCATCGCAACAGGATTGACAGGCATC 1979
650 rIleHisSerGlnGluMetLeuGlyGlnPheLeuGlnGluValSerValM 667
1980 CATTACAGCCAGGAGATGCTGGCGCAGTCTCTCCAGAGGTCCTCGTCA 2029
667 etAlaValMetLysThr 672
2030 TGGCAGTGATGAACCC 2046

seq_name: gb_pr:AF089094

seq_documentation_block:

LOCUS AF089094 2455 bp mRNA linear PRI 23-JAN-2001
DEFINITION Homo sapiens calpain-like protease CAPN10f mRNA, complete cds.
ACCESSION AF089094
VERSION AF089094.1 GI:10503947
KEYWORDS
SOURCE human.

ORGANISM Homo sapiens

REFERENCE
AUTHORS

Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus
Nat. Genet. 26 (2), 163-175 (2000)
2 (bases 1 to 2455)
Horikawa, Y. and Bell, G.I.
Direct Submission
Submitted (02-Sep-1998) Howard Hughes Medical Institute, The University of Chicago, 5841 S. Maryland Avenue, MC1028, Chicago, IL

TITLE

JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL

60637, USA

FEATURES Location/Qualifiers

Source

1. .2455

/organism="Homo sapiens"

/db_xref="taxon:9606"

CDS

42..866

/note="type F; alternatively spliced"

/codon_start=1

/product="calpain-like protease CAPN10f"

/protein_id="A017972.1"

/db_xref="gi:10503948"

/translation="MRAGKATPARELFNDAPPAADSLFCDLSTPLAOFREDITWR
RPEICATPRLFPDPRGQVKGILGDCWFLCACAALQSRLLDQVLPQGPSPVAD
OERYGSETRIMQFRMVEYTDRLPLCLAGRCFSRCOREDFWMLPLEKGFVTLRA
PVGRAGGCPGCFDRRGRKMEPEGSRKRRPAGCARPLGAODLSAAPPEGVSDOL
LRAPQSRCPGAGVPCFLHCLGPAAGSGGVPVHPAADPEPLGPAVLALERGC"

BASE COUNT 429 a 769 c 827 g 430 t

ORIGIN

alignment_scores:

Quality: 3604.00

Length: 673

Ratio: 5.395

Gaps: 2

Percent Similarity: 99.257

Percent Identity: 99.257

alignment_block:

US-09-768-877-2 x AF089094

Align seg 1/1 to: AF089094 from: 1 to: 2455

```

1 MetcrgalaglyargglyalathProAlaargluleupheargaspal 17
|||||
42 ATGCGGCGGGCGCGGGCGCGACCCGCGCGAGAGCTGCCGGGAGCG 91
|||||
17 aAlaPheProAlaAlaAspSerSerleupheCysAspleuserhProL 34
|||||
92 CGCCTCCCGCCCGCGGAGCTCTGCTCTGCGACTGTCTAGCGCGC 141
|||||
34 euaAlaGlnPhearggluasplleThrTrpArgArpProGlnGluIleCys 50
|||||
142 TGGCCAGTTCGCGAGGACATCACTGAGAGCGGCCCGCCAGAGATTGT 191
|||||
51 AlaThrProArgleupheProAspAspProArggluGlnValLysGI 67
|||||
192 GCCACACCCCGCGCTGTTCCAGATGACCCAGGAGGAGGTGAAGCA 241
|||||
67 nglyleuleuleuglyaspcystrpheleucysAlaCysAlaAlaLeuGln 84
|||||
242 GGGGCTCTGGGGGATTGCTGTGCTGTGCTGCGCCGCGCTGCAGCA 291
|||||
84 ySserATGHSleuleuAspGlnValIleProProGlnProSerTrp 100
|||||
292 AGAGCAGGACACTCTGACACAGGTCACTCTCCGGGACAGCCAGGCTGG 341
|||||
101 AlaAspGlnGluTrpArgGlySerPheThrCysArgIleTrpGlnPheGI 117
|||||
342 GCCACACAGGAGTACCGGGGCTCTTCACTGCGCATTTGGCAGTTTGG 391
|||||
117 yArgTrpValGluValIleThrAspAspArgleuProCysLeuAlaGly 134
|||||
392 ACCTGGGTGGAGGTGACACAGATGACCGGCTCTGCTGCTGCAGGGA 441
|||||
134 rgleuCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
|||||
442 GACTCTTTTCTCCGCTGCCAGAGGAGATGTTCTGCTCCCTTA 491
|||||
151 LeuGlnLysValTyralAlaLysValHisGlySerTyrgluHisLeuTrp 167
|||||
492 CTGGA.....AAGGTCATGGGTCTTACGAGACCTGAGG 529
|||||
167 laGlyGlnValAlaAspAlaLeuValAspleuthrGlyGlyLeuAlaGlu 183
|||||
530 CGGGGAGGTGGGATGCCCTGTGATCCTGACCGGCGGCTGCGAGAA 579

```

```

184 ArgTrpAsnLeuLysGlyValAlaGlySerGlyGlnGlnAspArgPr 200
|||||
580 AGATGGACCTGAAAGCGGTAGCAGAAAGCGGACCGACGAGCAGCGC 629
|||||
200 oGlyArgTrpGlnHisArgTrpCysArgGlnLeuLeuHisLeuLysAsp 217
|||||
630 AGGCGCTGGAGACACAGACTTGTGCGCACCTGTCCACTGAAGAGCC 679
|||||
217 IncysleuIleSerCysCysValleuserProArgAlaGlyAlaArgGlu 233
|||||
680 AGTCTGATCAGCTGCTGCTGCTCAACCCACAGCAGCTGCCGGGAG 729
|||||
234 LeuGlyGlnPheHisAlaPheIleValSerAspleuArgGluGlnGln 250
|||||
730 CTGGGGAGTTCATGCTCTTATTTGTCGGAACCTGGGAGCTCCAGG 779
|||||
250 yGlnAlaGlyIncysIleLeuLeuLeuArgIleGlnAspProTrpGly 267
|||||
780 TCAGGCGGGCCAGTGCATCTGCTGCGGATCCAGAACCCCGGGGCC 829
|||||
267 rArgGlyTrpGlnGlyLeuTrpArgGluGlyGlyGlyTrpSerGln 283
|||||
830 GCGGCTGCTGGCAGGGGCTCTGAGAGAGGGGGGTGAAGCTGGAGCGAG 879
|||||
284 ValAspAlaAlaValAlaSerGluLeuLeuSerGlnLeuGlnGluGly 300
|||||
880 GTAGTGCAGCGGTAGCATCTGACTCTCTGCTCCAGCTCCAGAAAGGGA 929
|||||
300 uPheTrpValGluGluGluGluPheLeuArgGluPheAspGluLeuThr 317
|||||
930 GTTCTGGGTGAGAGAGAGAGATTCTCAGAGGATTTGAGACGCTCAAC 979
|||||
317 aIGlyTrpProValThrGlnAlaGlyHisleuGlnSerLeuTrpThrGlu 333
|||||
980 TTGGCTACCCCGTCCAGAGAGCGCGCCACCTGCAAGAGCTCTTACACAGAG 1029
|||||
334 ArgLeuLeuCysHisThrArgAlaLeuProGlnValAlaTrpValLysGI 350
|||||
1030 AGGCTGCTCTCCATACCGCGGGCGCTGCTGGGCTGGGCTCAAGGCGCA 1079
|||||
350 nSerAlaGlyGlyCysArgAsnAsnSerGlyPheProSerAsnProLys 367
|||||
1080 GTCAAGAGGAGGCTGCCGGAACAACAGCGGCTTCCACAGAACCCCAAT 1129
|||||
367 heTrpLeuArgValSerGluProSerGluValTyrlleAlaValleuGln 383
|||||
1130 TCTGCTGCGGGGCTCAGAAACCGAGTAGGCTACATTTGCCGCTCAG 1179
|||||
384 ArgSerArgLeuHisAlaAlaAspTrpAlaGlyArgAlaArgAlaLeu 400
|||||
1180 AGATCCAGGCTGCACGCGGGGACTGGGAGCGCGGCGGACAGTGGT 1229
|||||
400 lGlyAspSerHisThrSerTrpSerProAlaSerIleProGlyLysHis 417
|||||
1230 GGGGACAGTATACCTCTGAGACCCAGCAGAGATGCCGGGCAAGCACT 1279
|||||
417 yGlnAlaValGlyLeuHisleuTrpLysValGluLysArgArgValAsn 433
|||||
1280 ACCAGGCTGTGGGTCTCACTCTGAAAGTAGAAGCGGGGCTCAAT 1329
|||||
434 LeuProArgValleuserMetProProValAlaGlyThrAlaCysHisAl 450
|||||
1330 CTGCTTGAAGGCTCTGCTCATGCCCCCGGTGGTGCACCGCTCCATGC 1379
|||||
450 aTyAspArgGluValHisLeuArgCysGluLeuSerProGlyTyrl 467
|||||
1380 ATACGACCGGAGGTCCACCTGCTTGAAGCTTCAACGGGCTACATACC 1429
|||||
467 euaAlaValProSerThrPheLeuLysAspAlaProGlyGluPheLeuLeu 483
|||||
1430 TGGCTGTCCACAGACCTTCTGAAAGACGCGGAGGAGTTCTCTGCTC 1479

```


484	ArgValPheSerThrGlyArgValSerLeuSerAlaIleArgAlaValAl	500
1480	CGAGTCTTCTTACCGGGCGAGTCTCCCTTAGCGCCATCAGGCGAGTGC	1529
500	aTysAsnThrThrProGlyAlaAlaLeuProAlaGlyGlnThrGlyThrV	517
1530	CAAGAACACACCCCCGGGGCAGCCCTGCCCTGCCGGGAGTGGGGGACCG	1579
517	alGlnLeuArgGlySerTrpArgValGlyGlnThrAlaGlyGlySerArg	533
1580	TGCAGCTACGGGGTCTTGGAGAGTCGGCCACAGCGGGGGGCGAGCAGG	1629
534	AsnPheAlaSerTyrProThrAsnProCysPheProPheSerValProG	550
1630	ACTTTGCCCTCATACCCACCAACCCCTGCCTTCCCTTCCTCGGTCGCCGA	1679
550	uGlyProGlyProArgCysValArgIleThrLeuHisGlnHisCysArgP	567
1680	GGGCCCTGGCCCCCGCTGCGTCGCGATCACATCTGCATCAGCAGCTCGCGC	1729
567	roSerAspThrGluPheHisProIleGlyPheHisIlePheGlnValPro	583
1730	CCAGTGCACACGAGTTCCACCCCATCGGCTTCCATATCTTCCAGGTCCCA	1779
584	GluGlyGlyArgSerGlnAspAlaProPheLeuLeuGlnGluProLe	600
1780	GAGGTGGAGGAGCCAGAGCACCCCCACTGCTGCTGCAGAGAGCGCT	1829
600	uLeuSerCysValProHisArgTyrAlaGlnGluValSerArgLeuCysL	617
1830	GCTGAGCTGCGTGCCACATCGCTACGCCAGAGAGTGCAGCGGCTCTGCC	1879
617	eULeuProAlaGlyThrTyrLysValValProSerThrTyrLeuProAsp	633
1880	TCTTGCCTGCAGGCACCTACAGGTGTGCCCTCCACTACCTGCCGGAC	1929
634	ThrGluGlyAlaPheThrValThrIleAlaThrArgIleAspArgProse	650
1930	ACAGAGGGGGCTTCACACAGTCACCATGCCAACCAAGGATTGCAGAGGCATC	1979
650	rIleHisSerGlnGluMetLeuGlyGlnPheLeuGlnGluValSerValM	667
1980	CATTTCACAGCCAGGAGATCTGGGCCAGTTCCTCCAAGAGGTCTCGTCA	2029
667	etalavalMetLysThr	672
2030	TGGCAGTGATGAAAC	2046

seq_name: qb_pr:AB072744

seq_documentation_block:
LOCUS AB072744 2602 bp mRNA linear PRI 11-OCT-2001
DEFINITION Macaca fascicularis testis cDNA clone:QtsA-15844, full insert
sequence.
ACCESSION AB072744
VERSION AB072744.1 GI:16041091
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Macaca fascicularis adult male testis cDNA to mRNA,
clone_lib:macaque testis cDNA library QtsA clone:QtsA-15844.
ORGANISM Macaca fascicularis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
Cercopithecinae; Macaca.
REFERENCE 1 (bases 1 to 2602)
Hashimoto,K., Osada,N., Hida,M., Kusuda,J., Tanuma,R., Hirai,M.,
Terao,K. and Sugano,S.
Isolation of novel full-length cDNA clones from macaque testis cDNA

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2602)
AUTHORS Hashimoto K., Osada N., Hida M., Kusuda, J. and Sugano, S.
TITLE Direct Submission
JOURNAL Submitted (09-OCT-2001) Katsuyuki Hashimoto, National Institute of

Infectious Diseases, Division of Genetic Resources; 23-1, Toyama
 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan
 (E-mail: khashi@nih.go.jp, URL: <http://www.nih.go.jp/yoken/genebank/>;
 Tel: 81-3-5285-1111(ex.2120), Fax: 81-3-5285-1181)
 Lab host: TOP10
 Vector: pME18S-FL3 (Acc.No. AB009864)
 R. Site1: DraIII (CACTGTGTG)
 R. Site2: DraIII (CACCATGTG)
 Description: 1st strand cDNA was primed with an oligo(dT) primer
 [ATGTGGCTTTTCTTTTCTTTTCTTTT]; double-stranded cDNA was synthesized
 using specific 5' and 3' primers and amplified by PCR. The PCR
 product was digested with SfiI and size selection was performed to
 exclude fragments <1.5kb. The SfiI-digested PCR product was cloned
 into distinct DraIII sites of pME18S-FL3. XhoI sites just outside
 the DraIII sites can be used to isolate the cDNA insert. Libraries
 were constructed by oligo-capping method
 (Sugano et al., University of Tokyo, Institute of Medical Science).
 Custom primer used for sequencing
 (5' end primer [CTTCTGCTCTAAAGCTGCG];
 3' end primer [CGCAGCTCAGCTCGAGCAC]).

FEATURES source

CDS

BASE COUNT	483 a	803 c	864 g	452 t
ORIGIN				

```
alignment_scores:
  Quality: 3388.50
  Ratio: 5.245
  Percent Similarity: 96.131
  Percent Identity: 93.452
  Length: 672
  Gaps: 1
```

```
alignment_block:
PS-09-768-877-2 y AR072744
```

Align seg 1/1 to: AB072744 from: 1 to: 2602

1	MetArgAlaGlyArgGlyAlaThrProAlaArgGluLeuPheArgAspAl	17
187	ATGCGGGCGGGCGGGGCGCGACGCCGCGAGGAGCTGTTCGGGACGC	236
17	aAlaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrProL	34
237	CGCTTCCCGCGCGGGGACTCTCGCTCTCTCGGACTTGTCTACGCCCG	286

34 euAlaGlnPheArgGluAspIleThrTrpArgArgProGlnGluIleCys 50
|||||
287 TGGCCCGAGTTCCGAGGACATCACGTGGAGCGGCCCCAGGAGATTGT 336
51 AlaThrProGluLeuPheProAspAspProArgGluGlyGlnValLysG1 67

```

337 GCCATCGCCCGGTTGTTCCAGATGACCCGACAGAGGCGAGTGAAGCA 386
67 nGlyLeuLeuGlyAspCysTrpPheLeuCysAlaCysAlaAlaLeuGln 84
387 GGGGCGCTGGGGGATTGCTGTTCTGTGTGCTGTGGCCGCTGCACA 436
84 ySerArghisLeuLeuAspGlnValIleProProGlyInProSerTrp 100
437 AGAGCAGGACCTCTCGAACAAGTATCTCTCCAGAGACGCAAGCTCG 486
101 AlaAspGlnGluTyrArgGlySerPheTrpCysArgIleTrpGlnPheG 117
487 GCCGACAGAGTACCAAGGCTCTTCACCTGCGATTGGCAGTTGG 536
117 yArgTrpValGluValThrThrAspAspArgLeuProCysLeuAlaGly 134
537 AGCTGGGTGGAGGTGACCAAGATGACCGCTGCGATGCTCGAGGA 586
134 rGLeuCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
587 GACTCTGTTCTCCGCTCCAGAGAGAGATGTCTGCGCTCCCTTA 636
151 LeuGlnLysValTyrAlaLysValHisGlySerTyrGlnHisLeuTrpAl 167
637 CTGGAAAGGCTACGCCCAAGGCTCCATGGGTCTACGAGACCTGGG 686
167 aGlyGlnValAlaAspAlaLeuValAspLeuThrGlyGlyLeuAlaGlu 184
687 AGGCAAGGTGGCGGATGCTCTGTGTGACTGTGCTGCTGCGAGAA 736
184 rGTrpAsnLeuLysGlyValAlaGlySerGlyGlnGlnAspArgPro 200
737 GATGAGCCTGAAAGGCGGTAGCAGAGAGGCGAGGCGACAGCAGGCTG 786
201 GlyArgTrpGlnHisArgThrCysArgGlnLeuLeuHisLeuLysAsp 217
787 GCGCCGTGGAGCACAGGACCTTGTGGGAGCTCCGCTGMAAGACA 836
217 nCysLeuIleSerCysCysValLeuSerProArgAlaGlyAlaArgGlu 234
837 GAGTCTCACTACACTGCTCGCTCTACGCCACGAGAGAGGTGCGGAG 886
234 euGlyLysPheHisAlaPheIleValSerAspLeuArgGluLeuGlnGly 250
887 TGGGGAGGTTCATGCTTCATTGCTCGAGACTCGGGAGCTGACAGAT 936
251 GlnAlaGlyGlnCysIleLeuLeuLeuArgIleGlnAsnProTrpGlyAr 267
937 CAGGGGGGCGACAGCATCTGCTGCGGATCCAGAACCCCTGGGGCGG 986
267 gArgCysTrpGlnGlyLeuTrpArgGluGlyGlyGlnGlyTrpSerGln 284
987 GCGGTGCTGGCAGGGGCTCTGGAGAGAGGGGTGAAGGTGGAGCCAG 1036
284 aLAspAlaAlaValAlaSerGluLeuLeuSerGlnLeuGlnGlyGlu 300
1037 TAGATGACAGCGGTACGTCTGAGCTCTGCTCCAGCTCCAGAGAGGAG 1086
301 PheTrpValGluGlnGluPheLeuArgGluPheAspGluLeuThrVa 317
1087 TTCGTGGGTGGAGAGAGAGTTCCTCAGGAGTTTGACGATACCAT 1136
317 lGlyTrpValThrGlnAlaGlyHisLeuGlnSerLeuTyrThrGln 334
1137 TGGCTACCAATCACAGAGGCGGCGACCTTCAGAGCTTACACAGGA 1186
334 rGLeuLeuCysHisThrArgAlaLeuProGlyAlaTrpValLysGlyGln 350
1187 AGCTGCTCTGGCACACTCGGGGCGCTGGGGCTGGGTCAAGGGGAG 1236
351 SerAlaGlyGlyCysArgAsnAsnSerGlyPheProSerAsnProLysPh 367
|||||
1237 TCAGCAGAGAGCTGCCGGAACAAGCGGCTTCCACAGAACCCCAATT 1286
367 eTrpLeuArgValSerGluProSerGluValTyrIleAlaValLeuGln 384
1287 CTGGCTGGGGGTCTGAGAACCGAGTACATTTCCTCCGCTGCAGAGA 1336
384 rGSerArgLeuHisAlaAspTrpAlaGlyArgAlaArgAlaLeuVal 400
1337 GATCAAGGCTGGGTCACTGAGTGGAGGCGGCGGCGGCACTGTGTG 1386
401 GlyAspSerHisThrSerTrpSerProAlaSerIleProGlyLysHisTy 417
1387 GGTGACAGTACATCTTCGTGGAGGCCACAGACATCCAGGACACATA 1436
417 rGlnAlaValGlyLeuHisLeuTrpLysValGluLysArgArgValAsn 434
1437 CCAGGCTGGGCTGCAACTCTGGAGAGTGGAGAACGCGGCGGTCAATC 1486
434 euProArgValLeuSerMetProProValAlaGlyThrAlaCysHisAla 450
1487 TGCCAGGGTCTTTCACAGCCCGCTGGCTGGAGTGCATGCTGCACACA 1536
451 TyrAspArgGluValHisLeuArgCysGluLeuSerProGlyTyrTyrLe 467
1537 TAGACCGGGAGGTCCACTGCTGGAGGCTTCACCGGGCTACTACT 1586
467 uAlaValProSerThrPheLeuLysAspAlaProGlyLysPheLeuLeu 484
1587 GCGTCTCCCAAGCACCTTCTGMAAGAGCGCCAGGGAGGATTCCTGCTCC 1636
484 rGValPheSerThrGlyArgValSerLeuSerAlaIleArgAlaValAla 500
1637 GAGTCTTCTTACCGGGAGTCTCTTACGGCATGAGGAGGTGGCC 1686
501 LysAsnThrThrProGlyAlaAlaLeuProAlaGlyLysTrpGlyThrVa 517
1687 AAGAAGCGCACCCCGACAGAGCCCTGCTGGGGGAGTGGGGAGCCGT 1736
517 lGlnLeuArgGlySerTrpArgValGlyGlnThrAlaGlyLysSerArg 534
1737 GCAGCTGGGGGCTCTCTGAGAGCGCGACAGCGCGGGGAGCAGACGA 1786
534 snPheAlaSerTyrProThrAsnProCysPheProPheSerValProGlu 550
1787 ACTTGGCTGTACCCCAACCAACCCCTCTTCCCTTCTCGGCTCCGAG 1836
551 GlyProGlyProArgCysValArgIleThrLeuHisGlnHisCysArgPr 567
1837 GGGCTGGCGCCCGGCTGGCGGATCACTGTGATCAAGCATGCGACACC 1886
567 oSerAspThrLysPheHisProIleGlyPheHisIlePheGlnValProG 584
1887 CAGGACACCCAGTTCACCCCATCGGCTTCATATCTTCAGAGTCCAG 1936
584 lGlyGlyArgSerGlnAspAlaProProLeuLeuLeuGlnGluProLeu 600
1937 AGGTGGAGGAGCGAGAGCGCGCCGCTGCTGCTGCTGAGAGCCGCTG 1986
601 LeuSerCysValProHisArgTyrAlaGlnGluValSerArgLeuCysLe 617
1987 CTGAGGCTCGGCGGCAATCGGTATGCCAGAGAGTGAACGACTGTGCT 2036
617 uLeuProAlaGlyThrTyrLysValValProSerThrTyrLeuProAsp 634
2037 CCTGCTTCCAGGACACTACAGGCTGTGCTTCACCTACCTGCGACAGA 2086
634 hrgLysAlaPheThrValThrIleAlaThrArgGlyIleAspArgProSer 650
2087 CAGAGGGGGCTTCACTGTGACCATGCGMACCAGATGAGAGGCTC... 2132
651 lIleHisSerGlnGluMetLeuGlyGlnPheLeuGlnGluValSerValMe 667
2133 .....TCCGTCAT 2140

```

667 talavalmetlythr 672
2141 GCGAGTGAATAAC 2156

seq_name: gb_pat:ARI53291

seq_documentation_block:

LOCUS ARI53291 2297 bp linear PAT 08-AUG-2001

DEFINITION Sequence 5 from patent US 6235481.

ACCESSION ARI53291

VERSION ARI53291.1 GI:15120823

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE

1 (bases 1 to 2297)

AUTHORS Horikawa, Y., Oda, N., Hanis, C.L., Bell, G.I. and Cox, N.J.

TITLE Polynucleotides encoding calpain 10

JOURNAL Patent: US 6235481-A 5 22-MAY-2001;

FEATURES

Location/Qualifiers

1..2297

/organism="unknown"

BASE COUNT 404 a 713 c 772 g 408 t

ORIGIN

alignment_scores:

Quality: 3290.00 Length: 672

Ratio: 5.350 Gaps: 1

Percent Similarity: 91.518 Percent Identity: 91.518

alignment_block:

US-09-768-877-2 x ARI53291

Align seg 1/1 to: ARI53291 from: 1 to: 2297

1 MetArgAlaGlyArgGlyAlaThrProAlaArgGluLeuPheArgAspAl 17

42 ATGCGGG 91

17 aAlaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrProL 34

92 CGCCTTCCCCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 141

34 euAlaGlnPheArgGluAspIleThrTrpArgArgProGlnGluIleCys 50

142 TGCGCCAGTTCGCGAGGACATCAGCTGGAGCGGGCCCGCAGGATTCT 191

51 AlaThrProArgLeuPheProAspAspProArgGluGlyGlnValIysG 67

192 GCCACACCCCGGCTGTTCCAGATGACCCAGGAGGGGAGGTGAAGCA 241

67 nGlyLeuLeuGlyAspCysTrpPheLeuCysAlaCysAlaAlaLeuGln 84

242 GGGGTGCTGGGGATTGCTGCTTCTGTGCTGCTGCTGCTGCTGCTG 291

84 ySerArgHisLeuLeuAspGlnValIleProProGlyGlnProSerTrp 100

292 AGAGCAGGCACCTCTCTGGACCAAGTCTATCTCCCGGAGCGAGCTGG 341

101 AlaAspGlnGluTyArgGlySerPheThrCysArgIleTrpGlnPheG 117

342 GCGGACAGAGTACCGGGGGCTCTTCCACCTGCTGCTGCTGCTGCTG 391

117 yArgTrpValGluValThrThrAspAspArgLeuProCysLeuAlaGly 134

392 ACGCTGGGTGGAGTGACACAGATGACCGCTGCTGCTGCTGCTGCTG 441

134 rGluCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150

442 GACTCTGTTCTCCCGCTGCCAGAGGAGGATGCTTCTGGCTCCCTTA 491

151 LeuGluLysValTyrAlaLysValHisGlySerTyrTyrGluHisLeuTrpAl 167
492 CTGGAAGAGGTCTACCCCAAGGTCCATGGGTCTACAGACACCTGTGGGC 541
167 aGlyGlnValAlaAspAlaLeuValAspLeuThrGlyGlyLeuAlaGluA 184
542 CGGGCAGGTGGGGATGCTCTGTGGACCTGACCGCGGCTGGCAGAAA 591
184 rGTrpAsnLeuLysGlyValAlaGlySerGlyGlyGlnGlnAspArgPro 200
592 GATGGAACCTGAAGGCGGTAGCAGGAAGCGGAGCGGCGGAGGAGGCA 641
201 GlyArgTrpGluHisArgThrCysArgGlnLeuLeuHisLeuLysAspG 217
642 GGGCGCTGGGAGCACAGGACTTGTGGCAGCTGCTCCACCTGAAGGACCA 691
217 nCysLeuIleSerCysValLeuSerProArgAlaGlyAlaArgGluL 234
692 GTGCTGTGATCAGCTGCTGCTGCTCAGCCCGCAGAGCAGGTGCCCGGAGC 741
234 euGlyGluPheHisAlaPheIleValSerAspLeuArgGluLeuGlnGly 250
742 TGGGGAGGTTCATGCTTCTGCGACCTGCGGAGCTCGGAGGTCCAGG 791
251 GlnAlaGlyGlnCysIleLeuLeuLeuArgIleGlnAsnProTrpGlyAr 267
792 CAGGGGGCGCAGTGCATCTGCTGCGGATCCAGAACCCCTGGGGCGG 841
267 gArgCysTrpGlnGlyLeuTrpArgGluGlyGlyGlyTrpSerGlnV 284
842 GCGGTGCTGGCAGGGGCTCTGGAGAGAGGGGGGTGAAGGTGGAGCCAGG 891
284 aAlaPheAlaAlaValAlaSerGluLeuLeuSerGlnLeuGlnGluGlu 300
892 TAGATGACGCGGTAGCATCTGAGCTCTGCTCCAGCTCCAGGAAGGGGAG 941
301 PheTrpValGluGluGluPheLeuArgGluPheAspGluLeuThrVa 317
942 TTCTGGGTGGAGGAGGAGGTTCTCAGGAGTTTGAGGACCTCACCGT 991
317 lGlyTrpProValThrGluAlaGlyHisLeuGlnSerLeuTrpGluA 334
992 TGGGTACCGGGTACGAGGAGGGGGGGGGGGGGGGGGGGGGGGGGGG 1041
334 rGLeuLeuCysHisThrArgAlaLeuProGlyValTrpValLysGlyGln 350
1042 GCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1091
351 SerAlaGlyGlyCysArgAsnAsnSerGlyPheProSerAsnProLysPh 367
1092 TCAGCAGGAGGCTGCCGGAACAACAGCGGCTTTCCAGCAACCCCAATT 1141
367 eTrpLeuArgValSerGluProSerGluValTyrIleAlaValLeuGlnA 384
1142 CTGGGTGCGGGTCTCAGAACCGAGTGAGGTGTACATTGCGCTCTGCAGA 1191
384 rGSerArgLeuHisAlaAlaAspTrpAlaGlyArgAlaArgAlaLeuVal 400
1192 GATCCAGGCTGCACGCGGAGGAGTGGGAGGCGGCGGCGGCGGCGGCTG 1241
401 GlyAspSerHisThrSerTrpSerProAlaSerIleProGlyLysHisTy 417
1242 GGTGACAGTCATCTCTGCGAGCGCAGCAGCAGCAGCAGCAGCAGCAG 1291
417 rGlnAlaValGlyLeuHisLeuTrpLysValGluLysArgValAsnL 434
1292 CCAGGCTGTGGGTCTGACACCTCTGGAAGGTAGAGAAGCGCGGGTCAATC 1341
434 euProArgValLeuSerMetProProValAlaGlyThrAlaCysHisAla 450
1342 TGCTTAGGGTCTCTCCATGCCCCCGTGGCTGGCAGCCCGCTGCCATGCA 1391
451 TyrAspArgGluValHisLeuArgCysGluLeuSerProGlyTyrTyrLe 467

```

|||||
1392 TACGACCGGAGGTCACCTCGTGTGAGCTCAGCCGCGCTACTACT 1441
467 uAlaValProSerThrPheLeuLysAspAlaProGlyGluPheLeu 484
|||||
1442 GCGCTTCCCGACCGCTTCTTGAAGACCGCGGAGGCTTCCGCTCC 1491
484 rGValPheSerThrGlyArgValSerLeuSerAlaIleArgAlaVal 500
|||||
1492 GAGCTTCTCTACCGGCGAGCTCCCTAG..... 1522
501 LysAsnThrThrProGlyAlaAlaLeuProAlaGlyLutPrgLys 517
1522 ..... 1522
517 LGlntLeuArgLysSerTrpArgValGlyGlnThrAlaGlySerArg 534
1522 ..... 1522
534 snPheAlaSerTyrProThrAsnProCysPheProPheSerValPro 550
1522 ..... 1522
551 GlyProGlyProArgCysValArgLleThrLeuHisGlnHisCysArg 567
|||||
1523 GGCCTGGCCCCCGCTGCGTCGACACTCTGCATCAGCACTGCGCGCC 1572
567 oSerAspThrGluPheHisProIleGlyPheHisIlePheGlnVal 584
|||||
1573 CAGTGACACCGAGTCCACCCCATCGCTTCCATATCTTCCAGGCTCC 1622
584 LuGlyGlyArgSerGlnAspAlaProProLeuLeuGlnGluProLeu 600
|||||
1623 AGGCTGAGAGAGCCAGGACGACCCCATCGCTGCTCAGAGAGCGCTG 1672
601 LeuSerCysValProHisArgTyrAlaGlnGluValSerArgLeuCys 617
|||||
1673 CTGAGCTGCGGCACATCGCTACGCCGAGAGGACCGCGCTCGCT 1722
617 uLeuProAlaGlyThrTyrLysValValProSerThrTyrLeuProAsp 634
|||||
1723 CCGCTGCTGAGGACACTACAGGTTGTGCTCCCTCAGCTGCGGAGCA 1772
634 hrGluGlyAlaPheThrValThrIleAlaThrArgIleAspArgPro 650
|||||
1773 CAGAGGGGCGCTTCAAGTGCATCGCAGCAGAGATTGACAGGCCATCC 1822
651 IleHisSerGlnGluMetLeuGlyGlnPheLeuGlnGluValSerVal 667
|||||
1823 ATTACAGCCGAGGAGATGCTGGCCAGTTCCTCAAGAGTCTCCGTCAT 1872
667 tAlaValMetLysThr 672
1873 GGCAGCTGATGAAAC 1888
seq_name: gb_dr:AF089090
seq_documentation_block:
LOCUS AF089090 2297 bp mRNA linear PRI 23-JAN-2001
DEFINITION Homo sapiens calpain-like protease CAPN10b mRNA, complete cds.
ACCESSION AF089090
VERSION AF089090.1 GI:10503938
KEYWORDS
SOURCE
ORGANISM Homo sapiens
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 2297)
REFERENCE
AUTHORS Horikawa,Y., Oda,N., Cox,N.J., Li,X., Orho-Melander,M., Hara,M.,
Hinokio,Y., Lindner,T.H., Mashima,H., Schwarz,P.E., del
Bosque-Plata,L., Horikawa,Y., Oda,Y., Yoshinuchi,I., Collila,S.,
Polonsky,K.S., Wei,S., Concanon,P., Iwasaki,N., Schuitze,J.,
Baler,L.J., Bogardus,C., Groop,L., Boerwinkle,E., Hanis,C.L. and

```

```

TITLE
Genetic variation in the gene encoding calpain-10 is associated
with type 2 diabetes mellitus
JOURNAL Nat. Genet. 26 (2), 163-175 (2000)
MEDLINE 20472315
REFERENCE 2 (bases 1 to 2297)
AUTHORS Horikawa,Y. and Bell,G.I.
TITLE Direct Submission
JOURNAL Submitted (02-SEP-1998) Howard Hughes Medical Institute, The
University of Chicago, 5841 S. Maryland Avenue, MC1028, Chicago, IL
60637, USA
FEATURES
source
1..2297
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2q37"
42..1676
/note="type B: alternatively spliced"
/codon_start=1
/product="calpain-like protease CAPN10b"
/protein_id="A017968.1"
/db_xref="GI:10503939"
/translation="MRAGKATFPARELFKDAAPADSSIFCDLSTPLAOFREDITWR
RPEICATPRLPDDPRGOVKGILGDMFLCAALQKSHLLDOVIIPROPSWAD
OYRSGFTCRIMQGRWVETTDRLPCLAGRCFRCQREDVFTPLPLEKTYARVHG
STEHLMAGQVADALVDLTGSLERMLKCVAGSGGQDPRGRHRTCHQLHLKQOC
LISCCVLSPRAGARELGEFHAFLVSDRLQAGQCLLLRLONPWGRRCWGLMRE
GGGGMQVDAVAASELSQLQEBEWEVEEELREFDELTVGYPTVEACHLSLYTER
LCHTRPALPGAWKQGSAGCRNNSGFPSNPKFVLVSESEYLIIVLGRSLHADM
AGRAALVDSHRSMSPASIPGKHQVAVGLHLMKVKRRVNLPRVLSMPVAGTACHA
YRREHLRCELSPGYLAIVSPFLKDAPEFLKRVSTGRVSLRALAPASISLCTST
ASPVPTSSRPSAISRSQREBARTHPCCSRRC"
BASE COUNT 404 a 713 c 772 g 408 t
ORIGIN
polya_signal
2289..2294
alignment_scores:
Quality: 3290.00 Length: 672
Ratio: 5.350 Gaps: 1
Percent Similarity: 91.518 Percent Identity: 91.518
alignment_block:
US-09-768-877-2 x AF089090 ..
Align seg 1/1 to: AF089090 from: 1 to: 2297
1 MetArgAlaGlyArgGlyAlaThrProAlaArgLysLeuPheArgAspAl 17
|||||
42 ATGCGGGCGGGCGGGCGGCGGCGGCGGAGGAGCTGTTCCGGGAGCG 91
|||||
17 aAlaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrPro 34
|||||
92 CGCCTCCCGCGCGCGGACCTCGCTCTCTGCGACTGTCAGTGGCGCGC 141
|||||
34 euAlaGlnPheArgGlyAspPheThrTrpArgArgProGlnGluIleCys 50
|||||
142 TGGCCGAGTTCGCGGAGACATCAGTGCAGGCGGCCCGCAGAGATTGCT 191
|||||
51 AlaThrProArgLysLeuPheProAspAspProArgGlyGlnValLysG 67
|||||
192 GCCACACCGCGCGCTGTTCCAGATGACCCACGGAAGGCGAGGTGAAGA 241
|||||
67 nglyLeuLeuGlyAspCysTrpPheLeuCysAlaCysAlaAlaLeuGln 84
|||||
242 GGGCTGCTGAGGGGATGCTGCTTCTGCTGCTGCGCGCGCTGCAGCA 291
|||||
84 ysserArgHisLeuLeuAspGlnValIleProProGlyGlnProSerTrp 100
|||||
292 AGACACAGGACATCTCTGAGACAGTCACTTCCGCGGACAGCGAGCTGG 341
|||||
101 AlaAspGlnGlyArgArgLysSerPheThrCysArgLleTrpGlnPheG 117

```

|||||
342 GCCACGAGGAGTACCGGGGCTCTTCCACCTGCGCATTTGGCAGTTTGG 391
|||||
117 YATrPValGluValThrThrAspArgLeuProCysLeuAlaGlyA 134
|||||
392 ACGCTGGTGGAGGTGACACAGATGACCGCTGCCGTGCTTGCAGGGA 441
|||||
134 rGLeuCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
|||||
442 GACTCTGTTCTCCCGTGCAGAGGAGGATGTGTCTGGCTCCCTTA 491
|||||
151 LeuGluLysValTyrAlaLysValHisGlySerTyrGluHisLeuTrpAl 167
|||||
492 CTGGAAGGTCACGCCCAAGGTCATATGGTCTTACGAGACCTGTGGGC 541
|||||
167 aGlyGlnValAlaAspAlaLeuValAspLeuThrGlyLeuAlaGluA 184
|||||
542 CGGCAGGTGGCGATGCCCTGTGTGGACCTGACCGCGGCTGGCAGAA 591
|||||
184 rGTrpAsnLeuLysGlyValAlaGlySerGlyGlnGlnAspArgPro 200
|||||
592 GATGGAACCTGAAGGCGGTAGCAGGAAGCGAGGCCAGCAGCAGGCCA 641
|||||
201 GlyArgTrpGluHisArgThrCysArgGlnLeuLeuHisLeuLysAspG 217
|||||
642 GGCCTGTGGAGCACAGGACTTGTGGCAGCTGTCCACCTGAAGGACCA 691
|||||
217 nCysLeuIleSerCysCysValLeuSerProArgAlaGlyAlaArgGluL 234
|||||
692 GTGCTGTATCAGCTGTGCTGTGCTAGCCCCCAGCAGAGGTGCCCGGAGC 741
|||||
234 euGlyGluPheHisAlaPheIleValSerAspLeuArgGluLeuGlnGly 250
|||||
742 TGGGGAGTTCCATGCTTCTTCGACCTGCGGGAGCTCCAGGGT 791
|||||
251 GlnAlaGlyGlnCysIleLeuLeuLeuArgIleGlnAsnProTrpGlyAr 267
|||||
792 CAGCGGGCCAGTGCATCTCTGCTGCTGCGGATCCAGAACCCCTGGGGCCG 841
|||||
267 aArGysTrpGlnGlyLeuTrpArgGluGlyGlyGlyTrpSerGlnV 284
|||||
842 GCGGTGCTGGCAGGGGCTCTGGAGAGAGGGGGGTGAAGGTGGAGCCAGG 891
|||||
284 alAspAlaValAlaSerGluLeuLeuSerGlnLeuGlnGluGlyGlu 300
|||||
892 TAGATGACGGGTAGCATCTGAGCTCTGCTCCAGCTCCAGGAAGGGAG 941
|||||
301 PheTrpValGluGluGluPheLeuArgGluPheAspCyluLeuThrVa 317
|||||
942 TTCCTGGGTGGAGGAGGAGGATTCCTCAGGGAGTTTGCAGGAGCTCACCGT 991
|||||
317 lGlyTyrProValThrGluAlaGlyHisLeuGlnSerLeuTyrThrGluA 334
|||||
992 TGCTACCCGGTCAAGGAGCGCGCCACCTGCGAGAGCTCTACACAGAGA 1041
|||||
334 rGLeuLeuCysHisThrArgAlaLeuProGlyAlaTrpValLysGlyGln 350
|||||
1042 GGCTGCTCTGCCATACGCGGGGCTGCTGGGGCTGGGTCAAGGGCCAG 1091
|||||
351 SerAlaGlyCysArgAsnAsnSerGlyPheProSerAsnProLysPh 367
|||||
1092 TCAGCAGGAGGCTGCCGGAACAACAGCGGCTTTCCAGCAACCCCAAT 1141
|||||
367 eTrpLeuArgValSerGluProSerGluValTyrIleAlaValLeuGlnA 384
|||||
1142 CTGGCTCGGGTCTCAGACCGAGTGGGTGTACATTGGCGTCTGCAGA 1191
|||||
384 rGSerArgLeuHisAlaAlaAspTrpAlaGlyArgAlaArgAlaLeuVal 400
|||||
1192 GATCCAGGCTGCACGGCGGAGTGGGCAGCGCGCGCGGACCTGGTG 1241
|||||
401 GlyAspSerHisThrSerTrpSerProAlaSerIleProGlyLysHisTy 417
|||||

1242 GGTGACAGTCATCTTCTGGAGCCCGAGCATCCCGGCAAGCACTA 1291
|||||
417 rGlnAlaValGlyLeuHisLeuTrpLysValGluLysArgArgValAsnL 434
|||||
1292 CAGGCTGTGGGTCTGCACCTCTTGAAGGTAGAGAAGCGCGGTCAATC 1341
|||||
434 euProArgValLeuSerMetProProValAlaGlyThrAlaCysHisAla 450
|||||
1342 TGCCTAGGTCCTGTCCATGCCCCCGTGGCTGGACCGCGGTCCATGCA 1391
|||||
451 TyrAspArgGluValHisLeuArgCysGluLeuSerProGlyTyrTrpLe 467
|||||
1392 TAGCAGCGGAGGTCCACCTGCTGTGAGCTCTCACCGGGCTACTACCT 1441
|||||
467 uAlaValProSerThrPheLeuLysAspAlaProGlyGluPheLeuLeuA 484
|||||
1442 GGTGTGCCCCAGCACCTTCTGAAGGACGCGCAGGGGAGTTCTCTGCTCC 1491
|||||
484 rGValPheSerThrGlyArgValSerLeuSerAlaIleArgAlaValAla 500
|||||
1492 GAGTCTTCTACCGGGCGAGTCTCCCTTAG 1522
|||||
501 LysAsnThrThrProGlyAlaAlaLeuProAlaGlyGluTrpGlyThrVa 517
|||||
1522 1522
|||||
517 lGlnLeuArgGlySerTrpArgValGlyGlnThrAlaGlyGlySerArgA 534
|||||
1522 1522
|||||
534 snPheAlaSerTyrProThrAsnProCysPheProPheSerValProGlu 550
|||||
1522 1522
|||||
551 GlyProGlyProArgCysValArgIleThrLeuHisGlnHisCysArgPr 567
|||||
1523 GGCCTTGGCCCCCGCTGCTGCGCATCTCTGCATCAGCACTGCGCGGC 1572
|||||
567 oSerAspThrGluPheHisProIleGlyPheHisIlePheGlnValProG 584
|||||
1573 CAGTGACACCGAGTTCCACCCCATCGGCTTCCATATCTTCCAGGTCCTCAG 1622
|||||
584 luGlyGlyArgSerGlnAspAlaProProLeuLeuLeuGlnProLeu 600
|||||
1623 AGGTTGAAGAGGAGCAGGAGCAGCCCCCACTGCTGCTGAGGAGCGCTG 1672
|||||
601 LeuSerCysValProHisArgTyrAlaGlnGluValSerArgLeuCysLe 617
|||||
1673 CTGAGCTGCTGCCACATCGCTACGCCAGGAGGTGAGCGGCTCTGCT 1722
|||||
617 uLeuProAlaGlyThrTyrLysValValProSerThrTyrLeuProAspT 634
|||||
1723 CCTGCTGCAGGCACCTACAGGTTGTGCCCTTCCACCTACCTGCCGGACA 1772
|||||
634 hrGluGlyAlaPheThrValThrIleAlaThrArgIleAspArgProSer 650
|||||
1773 CAGAGGGGCTTTCACATGACCATCGCAACCCAGGATTCAGAGGCCATCC 1822
|||||
651 IleHisSerGlnGluMetLeuGlyGlnPheLeuGlnGluValSerValMe 667
|||||
1823 ATTACAGCCAGGAGATGCTGGCGCAGTTCTCTCAAGAGGTCTCGCTCAT 1872
|||||
667 tAlaValMetLysThr 672
|||||
1873 GGCAGTGATGAANAACC 1888
|||||

seq_name: gb_pat:AR153294

seq_documentation_block:

LOCUS AR153294

DEFINITION Sequence 11 from patent US 6235481.

ACCESSION AR153294

VERSION AR153294.1 GI:15120826

2516 bp DNA

linear

PAT 08-AUG-2001

Fri Sep 20 10:47:47 2002

us-09-768-877-2.rge

Page 17

[illegible]

|||||
1542 GAGAGCGCGGTCTCAATCTGCTAGGCTCTGTCATGCCCGCGTGGC 1591
444 aGlyThrAlaCysHisAlaTyrAspArgGluValHisLeuArgCysGluL 461
|||||
1592 TGGCACCGGTCTCCATGATAGACCGGGAGGTCCACCTGCTGTGTGAGC 1641
461 euSerProGlyTyrTyrLeuAlaValProSerThrPheLeuLysAspAla 477
1642 TCTCACCGGTCTACTACCTGCTGCTCCACGACCTCTCTGAGGACGCG 1691
478 ProGlyGluPheLeuLeuArgValPheSerThrGlyArgValSerLeuSe 494
|||||
1692 CCAGGGAGTTCTCTCTCCGAGTCTCTCTACCGGGCGAGTCTCCCTTAG 1740
494 tAlaIleArgAlaValAlaLysAsnThrThrProGlyAlaAlaLeuProA 511
1741 1741
511 laGlyGluTyrPheThrValGlnLeuArgGlySerTrpArgValGlyGln 527
1741 1741
528 ThrAlaGlyGlySerArgAspPheAlaSerTyrProThrAsnProCysPh 544
1741 1741
544 eProPheSerValProGluGlyProGlyProArgCysValArgIleThrL 561
1742 GGCCTTGGCGCCCGCTGCTGCTCCGATCACTC 1772
561 euHisGlnHisCysArgProSerAspThrGluPheHisProIleGlyPhe 577
1773 TGCATCAGCACTCGCGGCCAGTGCACCGAGTTCCACCCCATCGGCTTC 1822
578 HisIlePheGlnValProGluGlyGlyArgSerGlnAspAlaProLe 594
1823 CATATCTTCCAGGTCCCGAGGGTGGAGGAGCCAGGACGCCACCCCACT 1872
594 uLeuLeuGlnGluProLeuLeuSerCysValProHisArgTyrAlaGlnG 611
1873 GCTGTGCAGGAGCGCTGCTGAGCTGGTGGCCATCGCTACGCCCAGG 1922
611 luValSerArgLeuCysLeuLeuProAlaGlyThrTyrLysValValPro 627
1923 AGGTGAGCGGCTGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1972
628 SerThrTyrLeuProAspThrGluClyAlaPheThrValThrIleAlaTh 644
1973 TCCACTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2022
644 rArgIleAspArgProSerIleHisSerGlnGluMetLeuGlyGlnPheL 661
2023 CAGGATTCACAGGCGCATCCATTCACAGCAGGAGATGCTGGGCCAGTTCC 2072
661 euGlnGluValSerValMetAlaValMetLysThr 672
2073 TCCAGAGGCTCTCGCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2107

seq_name: gb_pr:AF089093

seq_documentation_block:
LOCUS AF089093 2516 bp mRNA linear PRI 23-JAN-2001
DEFINITION Homo sapiens calpain-like protease CAPN10e mRNA, complete cds.
ACCESSION AF089093
VERSION AF089093.1 GI:10503945

KEYWORDS

SOURCE

ORGANISM

human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2516)
AUTHORS Horikawa, Y., Oda, N., Cox, N.J., Li, X., Orho-Melander, M., Hara, M.,

Hinokio, Y., Lindner, T.H., Mashima, H., Schwarz, P.E., del Bosque-Plata, L., Horikawa, Y., Oda, Y., Yoshiuchi, I., Colilla, S., Polonsky, K.S., Wei, S., Concannon, P., Iwasaki, N., Schuizer, J., Baier, L.J., Bogardus, C., Groop, L., Boerwinkle, E., Hanis, C.L. and Bell, G.I.
Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus
Nat. Genet. 26 (2), 163-175 (2000)
20472315
2 (bases 1 to 2516)
Horikawa, Y. and Bell, G.I.
Direct Submission
TITLE
Submitted (02-SEP-1998) Howard Hughes Medical Institute, The University of Chicago, 5841 S. Maryland Avenue, MC1028, Chicago, IL 60637, USA
FEATURES
source
1. .2516
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2q37"
42..1376
/note="type E; alternatively spliced"
with type 2 diabetes mellitus
/codon_start=1
/product="calpain-like protease CAPN10e"
/protein_id="AAG17971.1"
/db_xref="GI:10503946"
/translation="MRAGKATPARELFRDAAPPAADSLFCDLSTPLAQFREDITMR
RPEICATPRLEPDDPREGQVKGGLDCWFLCACAALQKSHLLDQVTPPGQSWAD
QYRGSFTCRIMQFGRVETDDRLPCLAGRLCFRCQREDFVFLPLLEKYAKVHG
SYEHLWAGQVADALVDLTGLAERWNLKGVAGSGQDQPRGWEHRTCOLLHLKDOC
LISCCVLPFRAGARELGEFPAFIVSDRLQAGOCILLRIQNPWGRPCWQGLWRE
GCGMSQVDAVASSELLSQLEGEFWEERLEFDELTVGYPTTEAGHLASLTER
LLCHTRALPGAWKVGQSGAGGCRNNSGFPNFRWLKVPSEVYITAVLQSRSLHAADM
AGRARALVGDSTHSWSPASIPGKHLQVQALHLMKGVTLGTLTFPPVPMWPT"

BASE COUNT 445 a 777 c 837 g 457 t
ORIGIN

alignment_scores:
Quality: 3239.50 Length: 745
Ratio: 5.267 Gaps: 2
Percent Similarity: 82.550 Percent Identity: 82.416

alignment_block:
US-09-768-877-2 x AF089093

Align seg 1/1 to: AF089093 from: 1 to: 2516

1 MetArgAlaGlyArgGlyAlaThrProAlaArgGluLeuPheArgAspAl 17
|||||
42 ATGCG 91
|||||
17 aAlaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrProL 34
|||||
92 CGCCTTCCCG 141
|||||
34 euAlaGlnPheArgGluAspIleThrTrpArgArgProGlnGluIleCys 50
|||||
142 TGGCCAGTTCGCGGAGGACATACGTTGAGGCGCGCGCGCGCGCGCGCGCG 191
|||||

51 AlaThrProArgLeuPheProAspAspProArgGluGlyGlnValLysG 67
|||||
192 GCCACACCCCGCGCTGTTTCCAGATGACCCACGCGGAGCGCGCGCGCGCG 241
|||||
67 nGlyLeuLeuGlyAspCysTrpPheLeuCysAlaCysAlaLeuGlnL 84
|||||
242 GGGGCTGCTGGGAGATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 291
|||||
84 ysSerArgHisLeuLeuAspGlnValIleProGlyGlnProSerTrp 100
|||||
292 AGAGCAGGACCTCTCTCGGACCGAGGTCATTCTCTCGGAGCAGCGCGCTGG 341
|||||

101 AlaAspGlnGluTyrArgGlySerPheThrCysArgIleTrpGlnPheG1 117
 |||||
 342 GCCGACCAAGAGTACCGGGGCTCTTCACTCTCCATTGGCACTTTGG 391
 |||||
 117 YAlGTrpAlaGluValThrThrAspAspArgLeuProCysLeuAlaGlyA 134
 |||||
 392 ACGGTGGGTGGAGGTACACAGATGACCGCTGGCCGTGGCAGAGGA 441
 |||||
 134 rGLeuCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
 |||||
 442 GACTCTGTTTCTCCGCTCCAGAGAGAGATGTTCTGGCTCCCTTA 491
 |||||
 151 LeuGluValValTyrAlaValHisGlySerTyrGlnHisLeuTrpAl 167
 |||||
 492 CTGGAAGGCTTACGCCAAGTCCATGGTCTTACGAGCACCTGGGCG 541
 |||||
 167 aGlyGlnValAlaAspAlaLeuValAspLeuThrGlyLeuAlaGluA 184
 |||||
 542 CGGGCAGGTGGCGGATGCCCTGGTGGACCTGACGGCGGCTGGCAGAA 591
 |||||
 184 rGTrpAsnLeuLysGlyValAlaLysSerGlyGlnGlnAspArgPro 200
 |||||
 592 GATGGAACCTGAAGGCGGTACAGAAACGGAGCCACAGACAGGCA 641
 |||||
 201 GlyArgTrpGlnHisArgThrCysArgGlnLeuLeuHisLeuLysAspG1 217
 |||||
 642 GCGCGCTGGAGCACAGGACTTGTGGCAGCTGCTCCACTGAAGAGCA 691
 |||||
 217 nCysLeuLysSerCysValLeuSerProArgAlaGlyAlaArgGlu 234
 |||||
 692 GTGTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 741
 |||||
 234 euGlyGlnPheHisAlaPheIleValSerAspLeuArgGlnLeuGlnGly 250
 |||||
 742 TGGGGAGATTCCATGCCCTTCACTTCTCGAGCTCGGGAGCTCCAGGT 791
 |||||
 251 GlnAlaGlnCysIleLeuLeuLeuArgIleGlnAsnProTrpGlyArg 267
 |||||
 792 CAGCGGGGCCAGTGCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 841
 |||||
 267 gArgCysTrpGlnGlyLeuTrpArgGlnGlyGlyGlyTrpSerGln 284
 |||||
 842 GCGGTGCTGGCAGGGGCTCTGGAGAGAGGGGTGAAGGTGAGCCAG 891
 |||||
 284 alAspAlaAlaValAlaSerGlnLeuLeuSerGlnLeuGlnGlnGly 300
 |||||
 892 TAGATGACGGGTGATCATCTGAGCTCTGCTCCAGCTCCAGAAAGGGAG 941
 |||||
 301 PheTrpValGlnGlnGlnPheLeuArgGlnPheAspGlnLeuTrpVal 317
 |||||
 942 TTTGGGTGGAGAGAGAGATTCTCAGGAGTTTACAGAGCTCACCGT 991
 |||||
 317 lGlyTyrProValThrGlnAlaGlnHisLeuGlnSerLeuTyrThrGlnA 334
 |||||
 992 TGGCTACCGGTACGAGGAGGGCCGACCTGCAAGGCTCTACACAGAGA 1041
 |||||
 334 rGLeuLeuCysHisThrArgAlaLeuProGlyAlaTrpValLysGln 350
 |||||
 1042 GCGTGGCTCTCCATACCGGGGCTGCTGGGCTGGGCTGAGAGGCGAG 1091
 |||||
 351 SerAlaGlyLysCysArgAsnAsnSerGlyPheProSerAsnProLysP 367
 |||||
 1092 TCAGCAGAGAGGTGCCGAACAGAGGCTTTCAGCAACCCCAAT 1141
 |||||
 367 eTrpLeuArgValSerGlnProSerGluValTyrIleAlaValLeuGlnA 384
 |||||
 1142 CTGGCTGGGGTCTCAAAACCGAGTGGTACTATGCGCTCTGAGAGA 1191
 |||||
 384 rGSerArgLeuHisAlaAlaAspTrpAlaGlyArgAlaArgAlaLeuVal 400
 |||||
 1192 GATCCAGGCTGCACGCGCGGACTGGGACAGGCGGGCGCTGGGCTG 1241
 |||||
 401 GlyAspSerHisThrSerTrpSerProAlaSerIleProGlyLysHisTy 417

|||||
 1242 GGTGACATCATCTTCTGTGAGGCCAGCCAGCATCCGGGCAAGCACTA 1291
 |||||
 417 rGlnAlaValAlaGlyLeuHisLeuTrpLys 426
 |||||
 1292 CCAGGCTGTGGGTCTGCACACTGTGGAAGGTTGTGACACTTGGCAACAC 1341
 |||||
 426 426
 1342 TGTTCCTGTCCCTTCATGATGTGGCCACATGATGTTCTTCTTCT 1391
 |||||
 426 426
 1392 GCMAAAGAGTTGCTGGAAGGCCACACTGTCCAGACGCCCCAGGTGCT 1441
 |||||
 426 426
 1442 GGGCCACGTCCTTTGTGGCCAGCTACAAAGAGACTTGCAGGCTGC 1491
 |||||
 427 Val 427
 1492 TGTCTGGACAGATPACTGGCCGAGGCCAAGTGAAGCCCGGATGGTA 1541
 |||||
 428 GluLysArgArgValAsnLeuProArgValLeuSerMetProProValAl 444
 |||||
 1542 GAGAAGCGCGGGTCAATCTCTCTAGGGTCTGTCCATGCCCCCGTGGC 1591
 |||||
 444 aGlyThrAlaCysHisAlaTyrAspArgGluValHisLeuArgCysGlu 461
 |||||
 1592 TGGCAGCGCGTGCATGCATGACGCGGAGAGTGCACCTGGTGTGAGC 1641
 |||||
 461 euSerProGlyTyrTyrLeuAlaValProSerThrPheLeuLysAspAla 477
 |||||
 1642 TCTACCGGGGTACTACTGGCTGCTCCACACCTTCTGTAAGAGAGCG 1691
 |||||
 478 ProGlyGlnPheLeuLeuArgValPheSerThrGlyArgValSerLeu 494
 |||||
 1692 CCAGGGAGTTCCTGCTCCGAGCTTCTCTACCGGGAGCTCTCTTAC 1740
 |||||
 494 rAlaIleArgAlaValAlaLysAsnThrThrProGlyAlaAlaLeuProA 511
 |||||
 1741 1741
 511 lGlyGlnTrpGlyThrValGlnLeuArgGlySerTrpArgValGlyGln 527
 |||||
 1741 1741
 528 ThrAlaGlyLysSerArgAsnPheAlaSerTyrProThrAsnProCysP 544
 |||||
 1741 1741
 544 eProPheSerValProGlyGlyProArgCysValArgIleThrL 561
 |||||
 1742 GGGCCTGGCCCCGGTGGGTGGGATCATCTC 1772
 |||||
 561 euHisGlnHisCysArgProSerAspThrGlnPheHisProIleGlyPhe 577
 |||||
 1773 TGCATCAGCACTGCGGCCAGTGACACCGAGTTCCACCCATCGGCTTC 1822
 |||||
 578 HisIlePheGlnValProGlnGlyLysArgSerGlnAspAlaProPhe 594
 |||||
 1823 CATATCTTCCAGGTCCCAAGGGTGGAGAGGACAGGACCCCACT 1872
 |||||
 594 uLeuLeuGlnGluProLeuLeuSerCysValProHisArgTyrAlaGln 611
 |||||
 1873 GCTGCTCAGAGAGCGGCTGCTGAGCTCGTCCACATCGGTACGCCAG 1922
 |||||
 611 lValSerArgLeuCysLeuLeuProAlaGlyThrTyrLysValValPro 637
 |||||
 1923 AGGTGACCGCGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1972
 |||||
 638 SerThrTyrLeuProAspThrGlnGlyAlaPheThrValThrIleAsn 644
 |||||

434 eupProAryValLeuSerMetProProValAlaGlyThrAlaCysHisAla 450
 1342 TCCCTAGAGGCTCTGTCATGACCCCGTGGCGACCCGCTGCATGCA 1391
 451 TyrAspArgGluValHisLeuArgCysGluLeuSerProGlyTyrTyr 467
 1392 TACGACCGGAGGCTCCACCTGCTGTGAGCTCTCACCGGGCTACTACT 1441
 467 uAlaValProSerThrPheLeuLysAspAlaProGlyGluPheLeu 484
 1442 GGCTGTCCCGACACCTCTCTAGAGACCCGCGAGGGAGTCTGCTCTC 1491
 484 rGValPheSerThrGlyArgValSerLeuSerAlaIleArgAlaValAla 500
 1492 GAGTCTTCTCTACCGGCGAGCTCTCCCTAG..... 1522
 501 LysAsnThrThrProGlyAlaAlaLeuProAlaGlyGluTyrGlyThrVa 517
 1522 1522
 517 IGlnLeuArgGlySerTyrArgValGlyGlnThrAlaGlyGlySerArg 534
 1522 1522
 534 snPheAlaSerTyrProThrAsnProCysPheProPheSerValProGlu 550
 1522 1522
 551 GlyProGlyProAlaGlyValArgIlePheLeuHisGlnHisCysArgPr 567
 1522 1522
 567 oSerAspThrGluPheHisProIleGlyPheHisIlePheGlnValProG 584
 1523 GTCCGAG 1529
 584 IuGlyGlyArgSerGlnAspAlaProProLeuLeuGlnGluProLeu 600
 1530 AGGCTGGAGAGGAGCCAGGACCCACCTGCTCTGCTGAGAGCCGCTG 1579
 601 LeuSerCysValProHisArgTyrAlaGlnGluValSerArgLeuCysLe 617
 1580 CTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1629
 617 uLeuProAlaGlyThrTyrLysValValProSerThrTyrLeuProAspT 634
 1530 CCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1679
 634 hrGluGlyAlaPheThrValThrIleAlaThrArgIleAspArgProSer 650
 1680 CAGAGGGGGCTTCCAGTGCATGCAACGACGAGATTGACAGGCAATCC 1729
 651 IleHisSerGlnGluMetLeuGlyGlnPheLeuGlnGluValSerValMe 667
 1730 ATTACAGCCAGGAGATGCTGGGCTGCTCTCCAGAGGCTCTCCGTCAT 1779
 667 tAlaValMetLysThr 672
 1780 GCGAGTATGATAAAC 1795

seq_name: gb-pr:AF089092

seq_documentation_block:

LOCUS AF089092

DEFINITION Homo sapiens calpain-like protease CAPN10d mRNA, complete cds.

ACCESSION AF089092

VERSION AF089092.1 GI:10503943

KEYWORDS

SOURCE

ORGANISM

human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

2204 bp mRNA linear PRI 23-JAN-2001

AF089092

GI:10503943

AF089092.1

AF089092.1

AF089092.1

AF089092.1

AF089092.1

AUTHORS

Horikawa, Y., Oda, N., Cox, N.J., Li, X., Orho-Melander, M., Hara, M.,
 Hinokio, Y., Lindner, T.H., Mashima, H., Schwarz, P.E., del
 Bosque-Plata, L., Horikawa, Y., Oda, Y., Yoshinuchi, I., Collilla, S.,
 Polonsky, K.S., Wei, S., Concanon, P., Iwasaki, N., Schülze, J.,
 Balier, L.J., Bogardus, C., Groop, L., Boerwinkle, E., Hanis, C.L. and
 Bell, G.I.

TITLE

Genetic variation in the gene encoding calpain-10 is associated
 with type 2 diabetes mellitus

Nat. Genet. 26 (2), 163-175 (2000)

2 (bases 1 to 2204)

Horikawa, Y. and Bell, G.I.

Submitted (02-SEP-1998) Howard Hughes Medical Institute, The
 University of Chicago, 5841 S. Maryland Avenue, MC1028, Chicago, IL
 60637, USA

FEATURES

source

1..2204
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="2"
 /map="2q37"
 42..1583
 /note="type D: alternatively spliced"
 /codon_start=1
 /product="calpain-like protease CAPN10d"
 /protein_id="AA017970.1"
 /db_xref="GI:10503944"
 /translation="MRAGCATPARELEFRDAPRADSSFLCDLSTPIAGFRDITMR
 RPOICATPRLFPDDPREGQVKGILDCDFLCAALAKSRHLLDQVTPPGPSNAD
 OYRGSTCTCINPGFQVETTTDRPLCLAGRCFCRQEDFVFLLEKVTAKVHG
 SYELHAGOVADLVDTGLARMLKGVAGSGGQDPRGRNHRCTRLILKQDC
 LISCVLSPVPAARELEFPAFTVSDRELQAGQOCILLRITONPWGRGCMGLNRE
 GGEWMSQVPAVAVSEILSOLQDGEFVVEEREFDELTVGYPTVEAGHLSLTER
 LICTRALPAMVKGSGAGCGRNNSGFPSPKFMRLSERSEYIVIAVLRSLAADW
 AGRAALVGDSSHMSWSPASIPGKHQAVGILHMKVKKRRVNPRLVTSMPVACTACHA
 YDREVLHRCLSGCVLLAVPSTFLKAPGFLRLRVSTGKRVSLRSQREKARTHPKCC
 CRRC"

CDS

BASE COUNT 390 a 672 c 752 g 390 t
 ORIGIN

alignment_scores:

Quality: 3090.50 Length: 672
 Ratio: 5.292 Gaps: 1
 Percent Similarity: 86.905 Percent Identity: 86.905

alignment_block:

us-09-768-877-2 x AF089092

Align seg 1/1 to: AF089092 from: 1 to: 2204

1 MetArgAlaGlyArgGlyAlaThrProAlaArgGluLeuPheArgAspAl 17
 42 ATGG 91
 17 AlaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrProL 34
 92 CGCCTTCCCGCCCGCGGAGCTCTGCTCTTCTGCGACTTGTCTACCGCCG 141
 34 euAlaGlnPheArgGluAspIleThrThrPargArgProGlnGlnIleCys 50
 142 TGGCCAGATTCCGAGAGCATCAGTGGAGCGGCGCCAGAGAGATTGTGT 191
 51 AlaThrProAlaGluPheProAspAspProArgGluGlnGlnValLysG 67
 192 GCCACACCGCGGCTCTTCCAGATGACCCAGGAGAGGCGAGTGAAGCA 241
 67 nGlyLeuGlnGlyAspCysTyrPheLeuCysAlaCysAlaAlaLeuGlnL 84
 242 GGGGCTGCTGGGGGAGTGTCTGCTCTGCTGCTGCTGCTGCTGCTGCA 291
 84 ySSerArgHisLeuLeuAspGlnValIleProProGlyGlnProSerTyr 100

|||||
292 AGACAGGACCTCTCGGACAGGTATCTCCGGACACCGAGCTGG 341
101 AlaAspGlnGluTyrArgGlySerPheThrCysArgIleTrpGlnPheG 117
|||||
342 GCCACACAGAGTACCGGGCTCTTCCACCTGTGCGCATTTGGCAGTTGG 391
117 YArgTrpValGluValThrThrAspAspArgLeuProCysLeuAlaGlyA 134
392 ACGCTGGGTGGAGGTGACACAGATGACCGCTGCCGTGCTTGCAGGGA 441
134 rGluCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
442 GACTCTGTTCTCCCGTCCAGAGGAGGATGTGTTCTGGCTCCCGCTTA 491
151 LeuGluLysValTyrAlaLysValHisGlySerTyrGluHisLeuTrpAl 167
492 CTGGAAGGTCTACGCCAAGGTCCATGGCTCTACGAGCACCTGTGGGC 541
167 aGlyGlnValAlaAspAlaLeuValAspLeuThrGlyGlyLeuAlaGlu 184
542 CGGGCAGGTGGCGATGCTGCTGGACCTTGACCGCGGCTGCGCAGAA 591
184 rGTrpAsnLeuLysGlyValAlaGlySerClyGlyGlnGlnAspArgPro 200
592 GATGGACCTGAAGGGCTAGCAGGAGGAGGAGGCCAGCAGGACAGGCCA 641
201 GlyArgTrpGluHisArgThrCysArgGlnLeuLeuHisLeuLysAspG 217
642 GGCGCTGGGAGCACAGGACTTCTGGCAGCTGCTCCACCTGAGGACCA 691
217 nCysLeuIleSerCysValLeuSerProArgAlaGlyAlaArgGluL 234
692 GTGCTGATCAGCTGTGCTGCTCAGCCCGCAGCAGGAGTCCCGGAGC 741
234 euGlyGluPheHisAlaPheIleValSerAspLeuArgGluLeuGlnGly 250
742 TGGGGAGTTCATGCTTCATTGTCTCGGACCTCGGGAGTCCAGGGT 791
251 GlnAlaGlyClnCysIleLeuLeuLeuArgIleGlnAsnProTrpGlyAr 267
792 CAGCGCGGCGAGTGCATCTGCTGCTGCGGATCCAGAACCCCTGGGGCGG 841
267 gArgCysTrpGlnGlyLeuTrpArgGluGlyGlyGlyGlyTrpSerGlnV 284
842 GCGGTCTGGCAGGGCTCTGGAGAGGGGGGTGAAGGTGGAGCCAGG 891
284 aLaspAlaValAlaValSerGluLeuLeuSerClnLeuGlnGluGlyGlu 300
892 TAGATCCAGCGGTAGCATCTGAGCTCTGTCCAGCTCCAGGAAGGGGAG 941
301 PheTrpValGluGluGluGluPheLeuArgGluPheAspGluLeuThrVa 317
942 TTCTGGGTGGAGGAGGAGGAGTTCCTCAGGGAGTTTGACGACTCACCGT 991
317 lGlyTrpProValThrGluAlaGlyHisLeuGlnSerLeuTyrThrGluA 334
992 TGCTACCCGGGTACAGGAGGGCGGCCACCTGCAGAGCCCTTACACAGAGA 1041
334 rGluLeuCysHisThrArgAlaLeuProGlyAlaTrpValLysGlyGln 350
1042 GCTGCTCTGCCATGCGGGGGCTGCTGCGGGGCTGGGTCAAGGGGCCAG 1091
351 SerAlaGlyClyCysArgAsnAsnSerGlyPheProSerAsnProLysph 367
1092 TCAGCAGGAGGCTGCCGGAACACAGGCGCTTCCAGCAACCCCAATT 1141
367 eTrpLeuArgValSerGluProSerGluValTyrIleAlaValLeuGlnA 384
1142 CTGGCTGCGGGTCTCAGAACCGAGTGAGGTGTACATTCGCTCCGTCACA 1191
384 rGSerArgLeuHisAlaAlaAspTrpAlaGlyArgAlaArgAlaLeuVal 400
|||||

1192 GATCCAGGCTGCACGGCGGACTGGGCGGCGGCGGCGGCGGCGGCTGGT 1241
401 GlyAspSerHisThrSerTrpSerProAlaSerIleProGlyLysHisTy 417
1242 GGTGACAGTCATATCTGTTGGAGCCAGCAGCATCCCGGCAAGCACTA 1291
417 rGlnAlaValGlyLeuHisLeuTrpLysValGluLysArgArgValAsnL 434
1292 CAGGGCTGTGGTCTGCACCTCTGGAAGGTAGAGACGCGGGGTCAATC 1341
434 euProArgValLeuSerMetProProValAlaGlyThrAlaCysHisAla 450
1342 TGCTAGGTGCTGTCCATGCCCGCGTGTGGCACCAGCGGTGCCATGCA 1391
451 TyrAspArgGluValHisLeuArgCysGluLeuSerProGlyTyrTyrLe 467
1392 TAGACCGGAGGTCCACCTGCTGTGTGAGCTCTCACCGGGCTACTACCT 1441
467 uAlaValProSerThrPheLeuLysAspAlaProGlyGluPheLeuLeuA 484
1442 GGCTGTCCCGCAGCACCTTCTGAGGAGCGCCAGGGGAGTTCTGCTCC 1491
484 rGValPheSerThrGlyArgValSerLeuSerAlaIleArgAlaValAla 500
1492 GAGTCTTCTCTACCGCGGAGTCTCCCTTAG 1522
501 LysAsnThrThrProGlyAlaAlaLeuProAlaGlyGluTrpGlyThrVa 517
1522 1522
517 lGlnLeuArgGlySerTrpArgValGlyGlnThrAlaGlyGlySerArgA 534
1522 1522
534 snPheAlaSerTyrProThrAsnProCysPheProPheSerValProGlu 550
1522 1522
551 GlyProGlyProArgCysValArgIleThrLeuHisGlnHisCysArgPr 567
1522 1522
567 oSerAspThrGluPheHisProIleGlyPheHisIlePheGlnValProG 584
1523 GTCCAG 1529
584 luGlyGlyArgSerGlnAspAlaProProLeuLeuLeuGlnGluProLeu 600
1530 AGGGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1579
601 LeuSerCysValProHisArgTyrAlaGlnGluValSerArgLeuCyste 617
1580 CTGAGCTGCTGCCATCGCTACGCGCCAGGAGGTGAGCGGCTCTGCT 1629
617 uLeuProAlaGlyThrTyrLysValValProSerThrTyrLeuProAspT 634
1630 CCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1679
634 hrGluGlyAlaPheThrValThrIleAlaThrArgIleAspArgProSer 650
1680 CAGAGGGGCGCTTCACAGTGACCATCGCAACAGGAGGATTGACAGGCATCC 1729
651 lIleHisSerGlnGluMetLeuGlyGlnPheLeuGlnGluValSerValMe 667
1730 ATTACAGCCAGGAGATGCTGGGCCAGTTCTTCCCAAGAGGTCTCGTCA 1779
667 tAlaValMetLysThr 672
1780 GGCAGTGATGAACACC 1795
seq_name: gb_ro:BC010969
seq_documentation_block:

LOCUS BC010969 2528 bp mRNA linear ROD 25-JUL-2001
 DEFINITION Mus musculus, calpain 10, clone MGC:13731 IMAGE:4159612, mRNA,
 complete cds.
 ACCESSION BC010969
 VERSION BC010969.1 GI:15012121
 KEYWORDS MGC.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 Strausberg, R.
 Direct Submission
 Submitted (23-JUL-2001) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 CONTACT: MGC help desk
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Jeffrey E. Green, M.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
 contact: amadansystemsbiology.org
 Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
 Greene, Mark Kettman and Anuradha Madan

REMARK
 COMMENT
 Contact: MGC help desk
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Jeffrey E. Green, M.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
 contact: amadansystemsbiology.org
 Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
 Greene, Mark Kettman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
 Series: IRK Plate: 18 Row: 1 Column: 9
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 6753259.

FEATURES

Source
 1.2528
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /db_xref="GI:15012122"
 /clone="MGC:13731 IMAGE:4159612"
 /tissue="type=Liver, normal. 5 month old male mouse."
 /clone_lib="NCI_CGAP_L19"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"
 111..2111
 /codon_start=1
 /product="calpain 10"
 /protein_id="AAH10969.1"
 /db_xref="GI:15012122"
 /translation="MRAVRAETPARELPFADAPASDSSILFNLSTPLAOFREDITMR
 RROEICATPOLFPDNDNGOVKQGLGDCMFLCAALQSOHLIDVPPROGSGSD
 OKYGFETCRIMQFGEVETIDRLPCLAGRLCSRCREDVETPLPLEKATAYVHG
 STEHLMAGOVADALVDLTGSLAEKRSKDYVTSAGQDDPSSGHNTRQQLHLKRC
 LISCVSLSRAGARELGEHFATISDLSQTSQGGIILLRHNWGRGQGLRE
 GEGHNQVPEAKESBELAQLOGEEFVEEERELREFDEVTIGYPTVEAHLISLTER
 VLCHRTLEGAWVTSOAGSCNNSCFNPFWLRLPESECVAVLQVPRRLDQ
 TRALGASPAVNPVPGKDOAVGLHMKVEKRKISLPRVLSAPVAGTCHAVDEPH
 LRCELSPGYVLLPSTFLKDVPGGLHVRFSKSTLSVRLATGASGTALPASEH
 ETVVQLOGCRRAQOTGAGSRNRPSTYRCPNCLPSPVSGACPRITRTILOCHRSIDSL
 HPIGHFVQVADGENDQSCSLIDEPPLSCVPHRAQDEVSRLCLLSTVGNTRYIPSTY
 LPDTGETFTVTLATRIDQSHISQEMLGQLQLOEVSFMAVMAK"

CDS

BASE COUNT 537 a 713 c 749 g 529 t
 ORIGIN

alignment_scores: Quality: 2969.00 Length: 671
 Ratio: 4.758 Gaps: 2
 Percent Similarity: 92.996 Percent Identity: 81.222

alignment_block:
 US-09-768-877-2 x BC010969

Align seg 1/1 to: BC010969 from: 1 to: 2528

```

1 MetatGAlaGlyArgGlyAlaThrProAlaArgGluLeuPheArgAspAl 17
111 ATGCGGGGGGTCGGGGCCGAGACGCCGGGGAGCTCTCCGGGAGCGC 160
17 aAlaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrProL 34
161 GGCATTCCTCCGCTCGACTCTCTCTTACACTTGTCCAGCCCTC 210
34 euAlaGlnPheArgGluAspIleThrThrParArgArgProGlnIleCys 50
211 TGGCCACGTTTCGGAGAGACATCATTTGGAGACGACCCAGAAATCTGT 260
51 AlaThrProArgLeuPheProAspAspProArgGluGlnValLysG 67
261 GCCACACCTCAGCTCTTCCAGATACCATGGAGGAGACAGGTGAACCA 310
67 nGlyLeuLeuGlyAspCysThrPheLeuCysAlaLysAlaLeuGln 84
311 AGGCTGCTGGAGATCTGCTGCTGTGCTGTGCTGCTGCTGCTGCTGCA 360
84 ySSerArgHisLeuLeuAspGlnValIleProProGlyGlnProSerTrp 100
361 AGAGTCACACCTCTCTGGACAGCTCTCCCTCCAGACAGCAGCAGCTGG 410
101 AlaAspGlnGluTrpArgGlySerPheThrCysArgIleTrpGlnPheG 117
411 TCTGACCAAAATACCAAGCTCTTCACTGCTGATTTGGCAATTGG 460
117 YArgTrpValGluValThrThrAspAspArgLeuProCysLeuAlaGly 134
461 ACACGCGAGAGAGACGACCATGATGATGCTGCTGCTGCTGCTGCGGA 510
134 rGluCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
511 GACTCTGCTTTCCTCCGCTGCGCAGACAGAGATGTTCTGCTGCTCCCTTA 560
151 LeuGluLysValIleValAlaLysValHisGlySerTrpGluHisLeuTrpAl 167
561 CTGCAAAAGCCTATGCTAAGCTCATGATGATGATGATGATGATGATGATG 610
167 aGlyGlnValAlaAspAlaLeuValAspLeuThrGlyGlyLeuAlaGlu 184
611 AGGCAATGCGACAGTCTCTTATGATCTCATCTGGAAGCCCTGGAGAA 660
184 rGTrpAsnLeuLysGlyValAlaGlySerGlyGlyGlnGlnAspArgPro 200
661 GGTGAGCTTGAAGAGTATTAAGAAAGCAGCGGCAGACAGACAGACACC 710
201 GlyArgTrpGluHisArgThrCysArgGlnLeuLeuHisLeuLysAspG 217
711 AGGTGTGGGAGACAGAACTGTGCGCAGCTACTCCACCTCAAGAGACCG 760
217 nCysLeuIleSerCysCysValIleSerProArgValAlaArgVal 234
761 GTGCTATATCAGCTGCTGCTGCTTATGACCCAGACAGAGGTCAGAGGAC 810
234 euGlyGluPheHisAlaPheIleValSerAspLeuArgGluLeuGlnGly 250
811 TCGAGAGATTCATGCTTATCATCTCAGATCTGACAGAGCTAGAGCT 860
251 GlnAlaGlyGlnCysIleLeuLeuLeuArgIleGlnAsnProTrpGlyAr 267
861 CAGACTGCGCAGGATATCTCTGCTGCGATTCACAAACCCCTGGGGCGC 910
267 gArgCysTrpGlnGlyLeuTrpArgGluGlyGlyGlnGlyTrpSerGln 284
911 GCGTGTGTTGGCAGGCGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
284 aAlaPalaAlaValAlaSerGluLeuLeuSerGlnLeuGlnGlyGlu 300
961 TAGAGCCAGCTAAGAGAGTCTGAGCTGCTGCGCAACTCAGAGAGAGAG 1010

```

301 PheTyrValGluGluGluPheLeuArgGluPheAspGluLeuThrVa 317
 1011 TTCTGGGTCGAGGAAGAGAGTTCTCTCAGGAGTTTGATGAGGTACCAT 1060
 317 lGlyTyrProValThrGluAlaGlyHisLeuGlnSerLeuTyrThrGlu 334
 1061 CGGCTACCCAGTCACAGAGCGCGGCCACCTACAGAGTCTCCACACAGA 1110
 334 rgLeuLeuCysHisThrArgAlaLeuProGlyAlaTyrValLysGlyGln 350
 1111 GGGTGTGTGCCATACCGCGGACACTCCCTGGTGGCTGGGTGACAGGGCAG 1160
 351 SerAlaGlyGlyCysArgAsnAsnSerGlyPheProSerAsnProLysPh 367
 1161 TCAGCAGGAGGCTGCGGGAACAACAGTTCTTCCCTGCAACCCCAAGTT 1210
 367 enrPLeuArgValSerGluProSerGluValTyrIleAlaValLeuGln 384
 1211 CTGGTTACGGCTCTTGGAAACCCAGCGAGGTGTGTGGCTTCTTCAGA 1260
 384 rgSerArgLeuHisAlaAlaAspTyrAlaGlyArgAlaArgAlaLeuVal 400
 1261 GACCCGGAGGCGCTTA.....GTGGGCCAGACTCGGGCAGCTGGCG 1301
 401 GlyAspSerHisThrSerTyrProAlaSerIleProGlyLysHisTy 417
 1302 GGT.....GCCAGTCTCCACCGGTGAACCTCCAGGCAAGACTA 1342
 417 rgAlaValGlyLeuHisLeuTyrLysValGluLysArgArgValAsnL 434
 1343 CCAGGCTGTGGCCCTGCACATCTGGAAGGTAGAGAAACGGAAGATCAGCC 1392
 434 euProArgValLeuSerMetProProValAlaGlyThrAlaCysHisAla 450
 1393 TGCCCCAGAGTCTGTGTCACCCCTGTGGCTGGCAGTGCATGCCATGCG 1442
 451 TyrAspArgGluValHisLeuArgCysGluLeuSerProGlyTyrTyrLe 467
 1443 TATGATCGTGAGATCCACTTGGTGTGTGAGCTCTCACCAGGCTACTACT 1492
 467 uAlaValProSerThrPheLeuLysAspAlaProGlyGluPheLeuLeu 484
 1493 GGCCGTCTCTAGCCTTTTGAAGGATGTGCCAGGGCAGTTCCTGTCTCA 1542
 484 rgValPheSerThrGlyArgValSerLeuSerAlaIleArgAlaValAla 500
 1543 GAGTCTTCTCCACTGGGAAATCTCCCTCAGTGGCGTCCAGGCTGGCCACC 1592
 501 LysAsnThrThrProGlyAlaAlaLeuProAlaGlyGluTyrPglyThrVa 517
 1593 AAGGTTGCATCGCTGGAACAGCCCTGCTGCAGCGAGTGGGAGACTGT 1642
 517 lGlnLeuArgGlySerTyrArgValGlyGlnThrAlaGlyGlySerArgA 534
 1643 GCAGTTGAGGGCTCTGGAGAGCTGGCCAGACAGCTGGGGGCGCAGGA 1692
 534 snPheAlaSerTyrProThrAsnProCysPheProPheSerValProGlu 550
 1693 ACTTTGGCTCTTACCCCTGCATCTCCCTGCTCCCTCTCTCTCTCTCTCT 1742
 551 GlyProGlyProArgCysValArgIleThrLeuHisGlnHisCysArgPr 567
 1743 GGTGTCGCCCCGCTACATCCGTATACCTACACCAACACTGCCGCT 1792
 567 oSerAspThrGluPheHisProIleGlyPheHisIlePheGlnValProG 584
 1793 CAGTCACAGCAGCTGCAACCCCATTTGGTTTCCATGTCTTTCAGGTTCCAG 1842
 584 lucGlyGlyArgSerGlnAspAlaProProLeuLeuGlnGluProLeu 600
 1843 CAGACGGTGCAGAACGAGCGGTGTTCCCTGCTGCTCCAGGAGCCACTG 1892

601 LeuSerCysValProHisArgTyrAlaGlnGluValSerArgLeuCysLe 617
 1893 CTAAGCTGTGTACCATCGCTACGCCAGGAAGTGAGCCGCTCTGCT 1942
 617 uLeuProAlaGlyThrTyrLysValValProSerThrTyrLeuProAspT 634
 1943 CCTTTCTGTGGGAACTACAGGATTTCTCCCTCCACCTACCTGCCAGATA 1992
 634 hrGluGlyAlaPheThrValThrIleAlaThrArgIleAspArgProSer 650
 1993 CAGAGGGTACCTTCACGGTAAACCATAGCAACCAAGATCGATAGGAGTCC 2042
 651 IleHisSerGlnGluMetLeuGlyGlnPheLeuGlnGluValSerValMe 667
 2043 ATCCACAGCCAGGAGATGCTGGCCAGCTGTCCAGAGGTCTCTCTTTAT 2092
 667 tAlaValMetLys 671
 2093 GGCAGTGTATGAAA 2105

CC secretory response or insulin mediated glucose transport.
XX
SQ Sequence 2620 BP; 448 A; 819 C; 899 G; 454 T; 0 other;

alignment_scores:
Quality: 3649.00 Length: 672
Ratio: 5.430 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-768-877-2 x AAA27476

Align seg 1/1 to: AAA27476 from: 1 to: 2620

1 MetArgAlaGlyArgGlyAlaThrProAlaArgGluLeuPheArgAspAl 17
178 ATCCGG 227
17 alaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrProL 34
228 CGCCTTCCCGCGCGCGGAGCTCTCGCTCTTCTCGGACTTGTCTACGCCG 277
34 euAlaGlnPheArgGluAspIleThrTrpArgArgProGlnGluIleCys 50
278 TGCCCCAGTTCCCGGAGGACATCACGTGGAGCGGCCCGCAGGAGATTGT 327
51 AlaThrProArgLeuPheProAspAspProArgGluGlyGlnVallysgl 67
328 GCCACACCCCGGCTGTTCCAGATGACCACCGGAGGCGGAGGTGAAGCA 377
67 nGlyLeuLeuGlyAspCysTrpPheLeuCysAlaCysAlaAlaLeuGlnL 84
378 GGGGCTGTGGGGGATGTTGTTCTCTGTGTCTGCTGCGCGCGCTGCAGA 427
84 ySerArgHisLeuLeuAspGlnValIleProProGlyGlnProSerTrp 100
428 AGACAGGCGACCTCTCTGGACACAGGTCTTCTCCGGGACGCGGAGCTGG 477
101 AlaAspGlnGluTyArgGlySerPheThrCysArgIleTrpGlnPhegl 117
478 GCGGACACAGGAGTACCGGGGCTCTTCACTGTGCGATTTGGCAGTTGG 527
117 yArgTrpValGluValThrThrAspAspArgLeuProCysLeuAlaGlyA 134
528 ACGCTGGGTGGAGGTGACACAGATGACCGCTGCGCTGCGCTTGCAGGGA 577
134 rgLeuCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
578 GACTCTGTTCTCCCGCTGCCAGAGGAGGATGTTCTGGCTCCCTTTA 627
151 LeuGluLysValTyAlaLysValHisGlySerTyArgGluHisLeuTrpAl 167
628 CTGGAAAAGGTCTACGCCAAGGTCTTACGGGTCTTACGAGCACCTGTGGC 677
167 aGlyGlnValAlaAspAlaLeuValAspLeuThrGlyGlyLeuAlaGluA 184
678 CGGGCAGGTGGGGAGTGGCTGGTGGACTGACCGGGCGGCTGGCAGAAA 727
184 rgTrpAsnLeuLysGlyValAlaGlySerGlyGlnGlnAspArgPro 200
728 GATGAACCTGAAGCGCTAGCAGAGGAGGAGGCGGAGGCGGAGGAGGCA 777
201 GlyArgTrpGluHisArgThrCysArgGlnLeuLeuHisLeuLysAspGl 217
778 GCGCCGCTGGGAGCACAGGACTTGTGGGAGCTGTCTCCACCTGAAGGACCA 827
217 nCysLeuIleSerCysValLeuSerProArgAlaGlyAlaArgGluL 234
828 GTGCTGTATCAGCTGCTGGTCTCAGCCCCAGAGCAGGTGCGCGGAGC 877
234 euGlyGluPheHisAlaPheIleValSerAspLeuArgGluLeuGlnGly 250
|||||

878 TGGGGGAGTTCCATGCTTCAATTGCTCGGACCTCGGGAGGCTCCAGGGT 927
251 GlnAlaGlyGlnCysIleLeuLeuLeuArgIleGlnAsnProTrpGlyAr 267
928 CAGCGGGGCCAGTGCATCTGCTGCTGCGGATCCAGAACCCCTCGGGCGC 977
267 gArgCysTrpGlnGlyLeuTrpArgGluGlyGlyGlyTrpSerGlnV 284
978 GCGGCTCTGGCAGGGGCTCTGGAGAGAGGGGGTGAAGGGTGGAGCCAG 1027
284 alaSpAlaAlaValAlaSerGluLeuLeuSerGlnLeuGlnGlyGlu 300
1028 TAGATCAGCGGTAGCATCTGAGCTCTGTCAGCTCCAGGAAGGGAG 1077
301 PheTrpValGluGluGluPheLeuArgGluPheAspGluLeuThrVa 317
1078 TTTCTGGGTGGAGGAGGAGGAGTTCTCAGGGAGTTTGACGAGCTCACCGT 1127
317 lGlyTyProValThrGluAlaGlyHisLeuGlnSerLeuTyThrGluA 334
1128 TGGCTACCCGGTCCAGGAGGCGGCCACCTGCAGAGCCTCTACACAGAGA 1177
334 rgLeuLeuCysHisThrArgAlaLeuProGlyAlaTrpValLysGlyGln 350
1178 GGCTGCTCTGCCATACGCGGGGCTGCTGGGCGCTGGGTCAAGGGCCAG 1227
351 SerAlaGlyCysArgAsnAsnSerGlyPheProSerAsnProLysph 367
1228 TCAGCAGGAGGCTGCCGGAACACAGCGGCTTTCCAGCAACCCCAAAAT 1277
367 eTrpLeuArgValSerGluProSerGluValTyIleAlaValLeuGlnA 384
1278 CTGGCTGGGGTCTCAGAACCCAGTGAGGTGTACATTGCCGTCTGCAGA 1327
384 rgSerArgLeuHisAlaAlaAspTrpAlaGlyArgAlaAlaArgAlaVal 400
1328 GATCCAGGCTGCACGCGCGGACTGGGAGCGCGCGGCGGCGGCTGGTG 1377
401 GlyAspSerHisThrSerTrpSerProAlaSerIleProGlyLysHisTy 417
1378 GGTGACAGTCATCTCTGTTGGAGCCAGCGGAGCATCCCGGGCAAGCACTA 1427
417 rGlnAlaValGlyLeuHisLeuTrpLysValGluLysArgArgValAsnL 434
1428 CCAGGCTGTGGGTCTGCACCTCTGGAAGGTAGAGAAGCGCGGCTCAATC 1477
434 euProArgValLeuSerMetProProValAlaGlyThrAlaCysHisAla 450
1478 TGCCCTAGGGTCTGTCCATGCCCGGCTGGCTGGCACCGCGTGCATGCA 1527
451 TyrAspArgGluValHisLeuArgCysGluLeuSerProGlyTyTrpLe 467
1528 TACGACCGGGAGGTCCACCTGGCTTGTGAGCTCTCACCGGGGTACTTACT 1577
467 uAlaValProSerThrPheLeuLysAspAlaProGlyGluPheLeuLeuA 484
1578 GGCTGCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1627
484 rgValPheSerThrGlyArgValSerLeuSerAlaIleArgAlaValAla 500
1628 GAGTCTTCTTACCGGGGAGTCTCCCTTAGCGGCATCAGGGGAGTGGCC 1677
501 LysAsnThrThrProGlyAlaAlaLeuProAlaGlyGluTrpGlyThrVa 517
1678 AAGAACACCCCGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1727
517 lGlnLeuArgGlySerTrpArgValGlyGlnThrAlaGlyGlySerArgA 534
1728 GCAGCTACGGGGTCTTGGAGAGTGGGCGCAGACGGCGGGGGGCGCAGAGA 1777
534 snPheAlaSerTyProThrAsnProCysPheProPheSerValProGlu 550
1778 ACTTTGCTCATACCCCAACCAACCCCTGCTTCCCTTCTCGGTCGCCCGAG 1827

134 rgLeuCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
|||||
549 GACTCTGTTTCTCCCGTGCAGAGGAGGATGTGTTCTGGCTCCCTTA 598
151 LeuGluLysValTyrAlaLysValHisGlySerTyrGluHisLeuTrpAl 167
|||||
599 CTGAAAGGCTTACGCCAAGGTCCATGGGTCTACGACGACCTGTGGGC 648
167 agLysGlnValAlaAspAlaLeuValAspLeuThrGlyGlyLeuAlaGluA 184
|||||
649 CGGGCAGGTGGCGATGCCCTGTGTGACCTGACCGCGGCGCTGGCAGAA 698
184 rgTrpAsnLeuLysGlyValAlaGlySerGlyGlyGlnGlnAspArgPro 200
699 GATGGAACCTGAAGGCGTAGCAGGAAGCGGAGCGCAGCAGGACGCGG 748
201 GlyArgTrpGluHisArgThrCysArgGlnLeuLeuHisLeuLysAspG 217
|||||
749 GGCCTGTGGAGCACAGGACTTGTTCGCCAGCTCTCCACCTGAAGGACCA 798
217 nCysLeuLysSerCysCysValLeuSerProArgAlaGlyAlaArgGluL 234
|||||
799 GTGTCTGATCAGCTGTGGTCTGCTCAGCCCGCAGAGAGGTGCCCGGAGC 848
234 euGlyGluPheHisAlaPheIleValSerAspLeuArgGluLeuGlnGly 250
849 TGGGGAGTTCATGCTTTCATTGTCTCGACCTGGCGGAGCTCCAGGCT 898
251 GlnAlaGlyGlnCystIleLeuLeuLeuArgIleGlnAsnProTrpGlyAr 267
|||||
899 CAGCGGGCGAGTGCATCTGCTGTGCGGATCCAGAACCTCTGGGGCGG 948
267 gArgCysTrpGlnGlyLeuTrpArgGluGlyGlyGlyTrpSerGlnV 284
|||||
949 CGCGTCTGGCAGGGCTCTGGAGAGAGGGGTGAAGGTGGAGCCAGG 998
284 aAlaSpAlaValAlaSerGluLeuLeuSerGlnLeuGlnGlyGlu 300
999 TAGATCAGCGGTAGCATCTGAGCTCTCTCCAGCTCCAGGAAGGGGAG 1048
301 PheTrpValGluGluGluPheLeuArgGluPheAspGluLeuThrVa 317
1049 TWTGGTGGAGAGAGAGGTTCCTCAGGGAGTTTGACGAGCTCACCGT 1098
317 lGlyTrpValThrGluAlaGlyHisLeuGlnSerLeuTyrThrGluA 334
1099 TGCTACCGCGTACAGAGCGCGCCACCTGCAGAGCTCTACACAGAGA 1148
334 rgLeuLeuCysHisThrArgAlaLeuProGlyAlaTrpValLysGlyGln 350
1149 GGTGCTCTGCCATACGCGGGCGCTGCTTGGGGCTGGGTCAAGGGCCAG 1198
351 SerAlaGlyGlyCysArgAsnAsnSerGlyPheProSerAsnProLysPh 367
1199 TCAGCAGGAGGCTGCCGGAAACACAGCGGCTTTCACAGAACCCCAAAAT 1248
367 eTrpLeuArgValSerGluProSerGluValTyrIleAlaValLeuGlnA 384
1249 CTGGCTGCGGGTCTCAGAACCGAGTGAGGTGTACATTGCCGTCTGCAGA 1298
384 rgSerArgLeuHisAlaAlaAspTrpAlaGlyArgAlaArgAlaLeuVal 400
1299 GATCCAGGTGACGCGGGCGACTGGGCAGCGCGCGCGGCACTGGGTG 1348
401 GlyAspSerHisThrSerTrpSerProAlaSerIleProGlyLysHisTy 417
1349 GGTGACAGTCATCTCTGGAGGCCAGCGAGCATCCCGGGCAAGCACTA 1398
417 rGlnAlaValGlyLeuHisLeuTrpLysValGluLysArgArgValAsnL 434
1399 CCAGGCTGTGGGTCTGCACCTCTGGAGGTAGAGAGCGCGGGTCAATC 1448

434 euProArgValLeuSerMetProProValAlaGlyThrAlaCysHisAla 450
|||||
1449 TGCTAGGCTCTGTCCATGCCCGCGTGGTGGCACCGGTCGCAATGCA 1498
451 TyrAspArgGluValHisLeuArgCysGluLeuSerProGlyTyrTrpLe 467
|||||
1499 TAGCAGCGGAGTCCACCTGTGCTGTGAGCTCTACCGGGCTACTACCT 1548
467 uAlaValProSerThrPheLeuLysAspAlaProGlyGluPheLeuLeuA 484
1549 GGTGTGCCCGCAGCACCTTCTGAAGGACGCGCAGGGGAGTTCTCTCTCC 1598
484 rgValPheSerThrGlyArgValSerLeuSerAlaIleArgAlaValAla 500
1599 GAGTCTTCTTACCGGGCGAGTCTCCCTTAGCGCCATCAGGCGAGTGGCC 1648
501 LysAsnThrThrProGlyAlaAlaLeuProAlaGlyGluTrpGlyThrVa 517
1649 AAGACACCGCGCGCGGCGAGCCCTGCTCGGGGGAGTGGGGGACCGT 1698
517 lGlnLeuArgGlySerTrpArgValGlyGlnThrAlaGlyGlySerArgA 534
1699 CGAGCTACGGGGTTCCTTGAGAGTGGCCAGACGCGGGGGCAGCAGCA 1748
534 snPheAlaSerTyrProThrAsnProCysPheProPheSerValProGlu 550
1749 ACTTGGCTCATACCCACCAACCCCTGCTTCCCTTCTCGGTCCCGAG 1798
551 GlyProGlyProArgCysValArgIleThrLeuHisGlnHisCysArgPr 567
1799 GGCCTTGGCGCGCGCTGCTCGGCATCACTTGCATCAGCACTGCGGCGC 1848
567 oSerAspThrGluPheHisProIleGlyPheHisIlePheGlnValProG 584
1849 CAGTGACACCGAGTTCACCCCTCGCTTCCATATCTTCCAGGTCCAG 1898
584 luGlyClyArgSerGlnAspAlaProProLeuLeuLeuGlnGluProLeu 600
1899 AGGTGGAAGGAGCCAGGACGACCCCACTGCTGTCGAGGCGCGCTG 1948
601 LeuSerCysValProHisArgTyrAlaGlnGluValSerArgLeuCysLe 617
1949 CTGAGCTGCGTGCACATCGCTACGCCAGAGGTGAGCGCGCTCTGCCT 1998
617 uLeuProAlaGlyThrTyrLysValValProSerThrTyrLeuProAspT 634
1999 CCTGCTGCGGGCACCTACAGGTGTGCTTCCCTCCACCTACCTGCCGGA 2048
634 hrGluGlyAlaPheThrValThrIleAlaThrArgIleAspArgProSer 650
2049 CAGAGGGGCGCTTTCACAGTGACCATCGCAACCAAGGATTGACAGGCCATCC 2098
651 lleHisSerGlnGluMetLeuGlyGlnPheLeuGlnGluValSerValMe 667
2099 ATTCAGCCAGGAGATGCTGGCGAGTTCCTTCCAAGAGGTCTCGCTCAT 2148
667 tAlaValMetLysThr 672
2149 GCAGTGATGAAACCC 2164
seq_name: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AAK94261
seq_documentation_block:
ID AAK94261 standard; cDNA: 2596 BP.
XX
AC AAK94261;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human full-length cDNA, SEQ ID NO: 2880.
XX
KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX

401 GlyAspSerHisThrSerTrpSerProAlaSerIleProGlyLysHisTy 417
1348 GGTGACAGTCATACCTTCGTGGAGCCAGCAGCATCCCGGCAAGCACTA 1397
417 rGlnAlaValGlyLeuHisLeuTrpLysValGlnLysArgValAsnL 434
1398 CCAGGCTGTGGTCTGCACCTCTGGAGGTAGAGAAGCGCGGGTCAATC 1447
434 euProArgValLeuSerMetProProValAlaGlyThrAlaCysHisAla 450
1448 TGCTAGGCTCTGTCTCATGCCGCCCGCTGGTGGCACCGCTGCCATGCA 1497
451 TyrAspArgGluValHisLeuArgCysGluLeuSerProGlyTyrTyrLe 467
1498 TAGACCGGGAGTCCACCTGCTGTGAGCTCTCACCGGGCTACTACCT 1547
467 uAlaValProSerThrPheLeuLysAspAlaProGlyGluPheLeuLeuA 484
1548 GGTGTCTCCACGACCTCTCTGAAGACGCGCCAGGGAGTTCTCTGCTCC 1597
484 rgValPheSerThrGlyArgValSerLeuSerAlaIleArgAlaValAla 500
1598 GAGTCTCTCTACCGGGCGAGTCTCCCTTAGCGCCATCAGGCGAGTGGCC 1647
501 LysAsnThrThrProGlyAlaAlaLeuProAlaGlyGluTrpGlyThrVa 517
1648 AAGAACACCGCCCGCGGGCAGCCCTGCTGCGGGGAGTGGGGACCGT 1697
517 lGlnLeuArgGlySerTrpArgValGlyGlnThrAlaGlyGlySerArgA 534
1698 GCAGCTACCGGGTCTCTGGAGAGTGGCCAGACGCGGGGGCAGCAGGA 1747
534 snPheAlaSerTyrProThrAsnProCysPheProPheSerValProGlu 550
1748 ACTTTGCTCATACCCACCCACCCCTGCTTCCCTCTCGCTGCCGAG 1797
551 GlyProGlyProArgCysValArgIlePheLeuHisGlnHisCysArgPr 567
1798 GGCCCTGGCCCGCGTGTGCTGCGCATCACTTGCATCAGCAGTGGCGGC 1847
567 oSerAspThrGluPheHisProIleGlyPheHisIlePheGlnValProG 584
1848 CAGTGACACCGAGTTCACCCCATCCGCTTCCATATCTTCCAGGTCCAG 1897
584 luGlyGlyArgSerGlnAspAlaProProLeuLeuGlnLeuGlnProLeu 600
1898 AGGTGGAGAGCAGCAGGAGCAGCCACCCACTCTCTGCTGAGAGCGCTG 1947
601 LeuSerCysValProHisArgTyrAlaGlnGluValSerArgLeuCysLe 617
1948 CTGAGCTGGTGCACATCGCTACGCCAGAGGTGAGCGGCTCTGCT 1997
617 uLeuProAlaGlyThrTyrLysValValProSerThrTyrLeuProAspT 634
1998 CTGCTGGGGACCTACAGGTGTGCTTCCCTTCCACTACCTTCCCGGACA 2047
634 hrGluGlyAlaPheThrValThrIleAlaThrArgIleAspArgProSer 650
2048 CAGAGGGGCTTACAGTGACCATCGACACCCAGGATGACAGGCGCATCC 2097
651 IleHisSerGlnGluMetLeuGlyGlnPheLeuGlnGluValSerValMe 667
2098 ATTCACAGCCAGAGATGCTGGGCGAGTCTCTCCAAAGAGGTCTCGTCTAT 2147
667 tAlaValMetLysThr 672
2148 GCAGTGATGAANACC 2163

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAA27481
seq_documentation_block:
ID AAA27481 standard; cDNA; 2455 BP.

XX AAA27481;
XX AC
XX DT 15-AUG-2000 (first entry)
XX DE Human calpain 10f cDNA.
XX DE
XX KW NIDDM1; non-insulin-dependent diabetes mellitus; CAPN10 gene;
XX KW calpain 10; calpain 10f; diapain-1; cysteine protease;
XX KW chromosome 2; human; diagnosis; therapy; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX CDS 42..866
XX FT /*tag= a
XX
XX WO200023603-A2.
XX PD 27-APR-2000.
XX PF 21-OCT-1999; 99WO-US24890.
XX PR 21-OCT-1998; 98US-0105052.
XX PR 13-MAY-1999; 99US-0134175.
XX PA (ARCH-) ARCH DEV CORP.
XX PI Polonsky KS, Horikawa Y, Oda N, Sreenan S, Zhou Y, Otani K;
XX PI Hanis CL, Bell GI;
XX DR WPI; 2000-339702/29.
XX DR P-PSDB; AAY79572.
XX PT Method for screening for type 2 diabetes mellitus comprises detecting a
XX PT polymorphism in a calpain encoding nucleic acid segment or a
XX PT protease-encoding nucleic acid segment -
XX PS Claim 65; Page 232-232; 257pp; English.
XX CC The present sequence is that of cDNA corresponding to a minor
XX CC transcript of the human calpain 10 gene, CAPN10 (see AA227475).
XX CC The sequence contains an open reading frame that encodes a protein
XX CC of 274 amino acids (see AAY79572), designated calpain 10f. The
XX CC CAPN10 gene consists of 15 exons spanning 32 kb. A complex pattern
XX CC of alternative splicing generates proteins of 672, 544, 517, 513,
XX CC 444, 274, 139 and 138 amino acids (see AAY79567-74), designated
XX CC calpain 10a to 10h, respectively. Calpain 10f is encoded by exons
XX CC 1-3*, 4-7 and 9-13 of the gene. The invention concerns the
XX CC identification of genes responsible for type 2 diabetes for use in
XX CC diagnostic and therapeutic applications. A G-to-A polymorphism in
XX CC intron 3 of the CAPN10 gene in the NIDDM1 region of chromosome 2
XX CC (UCSNP-43) shows evidence for linkage to type 2 diabetes. Claimed
XX CC methods for screening for a propensity for type 2 diabetes are
XX CC based on detection of a polymorphism in a calpain encoding nucleic
XX CC acid, especially UCSNP-10 of the CAPN10 gene. Methods are also
XX CC claimed for: producing calpain 10 polypeptides using calpain
XX CC 10-encoding polynucleotides, including the present sequence;
XX CC identifying modulators of calpain activity using calpain 10
XX CC polypeptides and nucleic acids; and using these modulators to
XX CC treat diabetes, in particular through the regulation of an insulin
XX CC secretory response or insulin mediated glucose transport.
XX SQ Sequence 2455 BP; 429 A; 769 C; 827 G; 430 T; 0 other;

alignment_scores:
Quality: 3604.00 Length: 673
Ratio: 5.395 Gaps: 2
Percent Similarity: 99.257 Percent Identity: 99.257
alignment_block:
US-09-768-877-2 x AAA27481

Align seq 1/1 to: AAA27481 from: 1 to: 2455

```
1 MetArgAlaGlyArgGlyAlaThrProAlaArgGluLeuPheArgAspAl 17
42 ATCGGGGGGGGGGGGGGGGGAGAGCCGGGAGGAGCTGTTCCGGAGCG 91
17 AlaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrPro 34
92 CGGCTTCCCGCGCGGAGCTCTGCTCTTCGACTTGTGTACGCCCG 141
34 euAlaGlnPheArgGluAspIleThrTrpArgArgProGlnIleCys 50
142 TGGCCCAAGTTCGCGAGAGATCAGTGAAGGGGCCCCAGAGGATTTG 191
51 AlaThrProArgLeuPheProAspAspProArgGluGlnValLysG 67
192 GCCACACCCCGGGCTTTCCAGATGACCCAGGGAGAGGAGGTGAGCA 241
67 nGlyLeuLeuGlyAspCysTrpPheLeuCysAlaCysAlaAlaLeuG 84
242 GGGCGCTGGGGGATGCTGTTCTGCTGCTGCTGCGCGCGCTGCACA 291
84 ySSerArgHisLeuLeuAspGlnValIleProProGlyGlnProSer 100
292 AGAGCAGACCTCTCGACACAGTCAATCTCCGGGACAGCCGAGCTGG 341
101 AlaAspGlnGluTrpArgGlySerPheThrCysArgIleTrpGlnPhe 117
342 GCCGACACAGAGTACCGGGGCTCCTTCACCTCTCCGACTTGGCATT 391
117 yArgTrpAlaGluValIleThrAspAspArgLeuProCysLeuAlaG 134
392 ACGCTGGGGAGGTGACACAGATGACCGCTGCGCTGCCTGCAGAGA 441
134 rGLeuCysPheSerArgCysGlnArgGluAspValPheTrpLeuPro 150
442 GACTGTGTTTCCTCCGCTGCCAGAGGAGATGTTCTGGCTCCCTTA 491
151 LeuGlnLysValTyrAlaLysValHisGlySerTyrGlnHisLeuTP 167
492 CTGAAA.....AAGGTCATGGGCTCAGACAGCACTGTGG 529
167 IagGlnValAlaAspAlaLeuValAspLeuThrGlyGlyLeuAlaG 183
530 CCGGGAGAGTGGGATGCTCTGTGAGCTGACCGGCGCTGCAGAA 579
184 ArgTrpAsnLeuLysGlyValAlaGlySerGlyGlnGlnAspArgPr 200
580 AGATGGAACTGAAAGGGGTGACAGAGAGGAGGAGGAGGAGGAGGAG 629
200 OGlyArgTrpGlnHisArgThrCysArgGlnLeuLeuHisLeuLysAs 217
630 AGCGCGCTGGAGACAGAGACTTTCGCGACCTGCTCACCCTGAAGACC 679
217 LncysLeuLeuSerCysCysValLeuSerProAlaGlyAlaArgGlu 233
680 AGTCTTGATAGTGTGCTGCTGCTCAGCCCAAGAGAGGAGGAGGAG 729
234 LeuGlyGluPheHisAlaPheIleValSerAspLeuArgGluLeuG 250
730 CTGGGGAGAGTTCATGCTTATGTTCTCGGACTGCGGAGCTCCAGGG 779
250 yGlnAlaGlyLncysIleLeuLeuLeuArgIleGlnAsnProTrpGly 267
780 TCAGGCGGGGAGTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 829
267 rGArgCysTrpGlnGlyLeuTrpArgGluGlyGlyGlyTrpSerGln 283
830 GGGGCTGTGGAGGGGCTCTGGAGAGAGGGGGTGAAGGGTGAAGCAG 879
284 ValAspAlaAlaValAlaSerGluLeuLeuSerLncysGlnGlyG 300
880 GTAGATCAGACGGGTAGCATCTGAGCTCTGTCCACTCCAGAGAGGA 929
300 uPheTrpValGluGluGluGluPheLeuArgGluPheAspGluLeuTr 317
930 GTTCTGGGTGGAGAGAGAGAGATTCCTCAGAGGATTGAGAGACTAC 979
317 aIGlyTrpProValIleThrGluAlaGlyHisLeuGlnSerLeuTyr 333
980 TTGGCTACCGGGTCAAGAGAGCGGCGCCTGACAGCTTACACAGAG 1029
334 ArgLeuLeuCysHisThrArgAlaLeuProGlyAlaIleTrpValL 350
1030 AGGCTGCTGTCCATACGGCGGGGCTGCTGGGGCTGGGGTCAAGGCA 1079
350 nSerAlaGlyGlyCysArgAsnAsnSerGlyPheProSerAsnProL 367
1080 GTACAGCAGAGGCTCGCCGAGACACAGAGGCTTCCACAGAACCCAA 1129
367 heTrpLeuArgValSerGluProSerGluValTyrIleAlaValLeu 383
1130 TCTGGCTGGGGTCTAGAACAGAGTGAAGTGTACATTCCCTGTGAG 1179
384 ArgSerArgLeuHisAlaAlaAspTrpAlaGlyArgAlaArgAla 400
1180 ACATCGAGGCTGCACGGCGGAGCTGGGACGCGGCGGCGGCACTGGT 1229
400 LgIlyAspSerHisThrSerTrpSerProAlaSerIleProGlyLys 417
1230 GGGTACAGTCACTCTGCGAGCCAGCAGACATCCCGGCAAGCACT 1279
417 yGlnAlaValGlyLeuHisLeuTrpLysValGlnLysArgValAsn 433
1280 ACCAGGCTGGGTGTGACCTCTGGAAGTGAAGAAGGGGGGTCAAT 1329
434 LeuProArgValLeuSerMetProProValAlaGlyThrAlaCysH 450
1330 CTGCTTGAAGGTCTCTCCATGCCCCCGGTGGACCGCTGCCATGC 1379
450 aTyrAspArgGluValHisLeuArgCysGluLeuSerProGlyTyr 467
1380 ATACAGCGGGAGGTCACTGCTGTGAGTCTACCGGGGCTACTACC 1429
467 euAlaValProSerThrPheLeuLysAspAlaProGlyGluPheLeu 483
1430 TGGCTGTCCCAAGCACTTCTGAAAGAGCGCCAGGGAGTCTGTGCT 1479
484 ArgValPheSerThrGlyArgValSerLeuSerAlaIleArgAlaVal 500
1480 CGAGCTTCTTACCGGGCGAGTCTTCCCTTACGCGCATCAGGAGCTG 1529
500 aLysAsnThrThrProGlyAlaAlaLeuProAlaGlyLutPrGlyThr 517
1530 CAAGAACAACACCCCGGGGAGGCTCTGCGGGGAGTGGGGAGACCG 1579
517 aGlnLeuArgGlySerTrpArgValGlyGlnThrAlaGlySerArg 533
1580 TGCAGCTACGGGGTCTTGGAGAGTCGGCCAGAGCGGGGGAGAGCG 1629
534 AsnPheAlaSerTyrProThrAsnProCysPheProPheSerValPro 550
1630 AACTTTGGCTCATACCCACCAACCCCGTCTCCCTTCTGGTCCCGA 1679
550 uGlyProGlyProArgCysValArgIleThrLeuHisGlnHisCysArg 567
1680 GGGGCTGGGCCCCCGCTGCGCTGCGCATCAGCTCAGTACAGCAGCG 1729
567 roSerAspThrGluPheHisProIleGlyPheHisIlePheGlnValPro 583
1730 CCACTGACACCGAGTTCACCCCAATCGGGTCCATCTTCCAGGTCCA 1779
584 GluGlyLysArgSerGlnAspAlaProProLeuLeuLeuGlnGluPro 600
1780 GAGGCTGAAGAGACCAAGAGACACCCCTGCTGCTGAGAGCGCT 1829
```


184 rgtprasnleuylsglyvalalaaglyserglyglyinglnaspargpro 200
 592 gatgaacctaagagggcgctgacagaaagcgagcgacagacagagcca 641
 201 glyatgttprglnhlsargthrCysArgGlnLeuLeuHlsLeuLysAspG1 217
 642 gggcgctggagacacagacacttctgacactgctccactgaaggacca 691
 217 ncysleuilesercyscysvalleuserproargalaaglyalarglul 234
 692 gtgtctgattcagctgctcgctgacgcccagacagagtgccggagac 741
 234 euglygluphehlsalaphelilevalseraspheulargluleuglly 250
 742 tggggagattccatgcttcattgtctgacgtccggagctccaggggt 791
 251 glnaaglyincysileleuleuleuargllieglnasnpotprg1yar 267
 792 cagcgggggcagtgcatctctgctgctggatccagaaacccctggggcg 841
 267 gtagcysrtprglnlyleutrtprargllyglyllyllytrpserglnv 284
 842 gcggtgtgaggggctctgagagaggggggtgaagggtgagagccag 891
 284 alaspalaialavalasergluleuleuserglnleugllyglu 300
 892 tagatgacagggagtagacatctgactctgctccagctccagaaaggag 941
 301 pheTPrValgluglulglupheleuarglupheaspgluleuthrva 317
 942 ttcctgggtgagagagagagattctctcagagagtttgaacagctcacc 991
 317 lgltytrprovalthrclualaglyhlsleuglnserleutythrslua 334
 992 tggcttaccggcgacagagggcgccacacgacgctctacacagaga 1041
 334 rgleuleucysHlsThrArgAlaleuProglialatrpvallysglyln 350
 1042 ggtgctctccatnaccggggcgctgctggggctggtgagagggcgag 1091
 351 serAlaglylcySarArgasnaasSerglyPheProSerAnProLysph 367
 1092 tcagcagagagctgcccgaacacagcggcttccacagcaaccccaatt 1141
 367 etrpheulargvalsergluprosergluvaltyrlllealavalleunla 384
 1142 ctggctggcggtctcagAACCGAGTGAAGTGTACATTGCCGCTCGAGA 1191
 384 rgserrgleuHlsAlaalaasptpralaglyarargalaalaleuval 400
 1192 gatccagcggcgacggcgagctggcgagcgccggcgagcactggtg 1241
 401 glyAspSerHlsThrSerTrpSerProAlaSerllleProgllyHsHsTy 417
 1242 ggtgacagatnacttctgtgagacccagacagatcccgggcagacata 1291
 417 rglinalavalglyleuHlsleutrpLysValgluLysArgArgValasn 434
 1292 ccagcgctggggtctgacacttgaaagtagaagacggcggtcacaac 1341
 434 euProArgValleusermetProProValAlaaglythrAlaCysHlsAla 450
 1342 tgccttagggctctgctcattgcccccggtgctggcaccgctgcacatga 1391
 451 tyraspargluValHlsleuargCysgluleuserProgllytyrlyle 467
 1392 ttagacacggagaggtccactgctttagcttccacgggctactgacct 1441
 467 ualavalProSerThrPheLeuLysAspAlaProgllyglupheleuLeu 484
 1442 ggcgtgccacacacacttctgaagagacggcagggaggttctctctgc 1491
 484 rgValPheSerThrArgValSerleuSerAlaileargalavalala 500

1492 GAGTCTTCTACCGGGGAGCTTCCTTAG..... 1522
 501 LysasnThrThrProglYAlaAlaLeuProAlaaglyutrtprgllythrva 517
 1522 1522
 517 lGlnleuargllysertrprargvalglylnthrAlaaglylseratga 534
 1522 1522
 534 snpheaLaserTyrProThrAnProCysPheProPheSerValProglu 550
 1522 1522
 551 GlyProglYProArgCysValarglleThrleuHlsGlnHlsCysArgPr 567
 1523 GGCCTGGCCCCCGCTGCTCCGATCATCTGCATCAGCACGTGGCGCC 1572
 567 oSerAspThcgluphehlsprollieglyPheHlsIlePheglntvalProG 584
 1573 CAGTGAACCGAGTTCCACCCCATGGCTTCATATCTTCAGAGTCCAG 1622
 584 lugiylgYArgSerGlnAspAlaProProleuleuleuglntProleu 600
 1623 AGGCTGGAAGAGCCAGAGCAGACCCACACGCTGCTGCAGAGACCGCTG 1672
 601 LeuSerCysValProHlsArgTyrAlaGlnlulvalSerArgleucysle 617
 1673 CTGAGTGTGCTGCCATCGCTACGCCAGAGAGGTGAGCGGCTGCTGCT 1722
 617 uLeuProAlaglyThrTyrlYsValValProSerThrTyrlleuProLyspr 634
 1723 Cctgcttgcagagcagcctacaaaggttgcttccacttaccctggcgaga 1772
 634 hrtglulglualapherThrValThrIleAlaThrArglleasparProser 650
 1773 CAGAGGGGGCCTTCACAGTGAACCATCCCAACAGAGATTGACAGCCATCC 1822
 651 lIleHlsSerGlnlumeleuglylnPheleuglnlulvalSerValme 667
 1823 ATTCCAGCCAGAGAGATGCTGGGCCAGTTCCTCCAAAGGTCTCTCCAT 1872
 667 talavalmetlysthr 672
 1873 GGCAGTGTATGAAAAACC 1888

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAA27480
 seq_documentation_block:
 ID AAA27480 standard; cDNA; 2516 BP.
 XX
 AC AAA27480;
 XX
 DT 15-AUG-2000 (first entry)
 XX
 DE Human calpain 10e cDNA.
 XX
 KW NIDDM1: non-insulin-dependent diabetes mellitus; CAPN10 gene;
 KW calpain 10; calpain 10e; diapain-1; cysteine protease;
 KW chromosome 2; human; diagnosis; therapy; ss.
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 42..1376
 FT /*tag= a
 XX
 PN WO200023603-A2.
 XX
 PD 27-APR-2000.
 XX
 PF 21-OCT-1999; 99WO-US24890.

XX 21-OCT-1998; 98US-0105052.
 PR 13-MAY-1999; 99US-0134175.
 XX (ARCH-) ARCH DEV CORP.
 XX Polonsky KS, Horikawa Y, Oda N, Sreenan S, Zhou Y, Otani K;
 PI Hanis CL, Bell GI;
 XX P-PSDB; AAY79571.
 DR WPI; 2000-339702/29.
 XX Method for screening for type 2 diabetes mellitus comprises detecting a
 PT polymorphism in a calpain encoding nucleic acid segment or a
 PT protease-encoding nucleic acid segment
 XX Claim 65; Page 229-230; 257pp; English.
 XX The present sequence is that of cDNA corresponding to a minor
 CC transcript of the human calpain 10 gene, CAPN10 (see AA27475).
 CC The sequence contains an open reading frame that encodes a protein
 CC of 444 amino acids (see AAY79571), designated calpain 10e. The
 CC CAPN10 gene consists of 15 exons spanning 32 kb. A complex pattern
 CC of alternative splicing generates proteins of 672, 544, 517, 513,
 CC 444, 274, 139 and 138 amino acids (see AAY79567-74), designated
 CC calpain 10a to 10h, respectively. Calpain 10e is encoded by exons
 CC 1-10* and 11-13 of the gene. The invention concerns the
 CC identification of genes responsible for type 2 diabetes for use in
 CC diagnostic and therapeutic applications. A G-to-A polymorphism in
 CC intron 3 of the CAPN10 gene in the NIDDM region of chromosome 2
 CC (UCSNP-43) shows evidence for linkage to type 2 diabetes. Claimed
 CC methods for screening for a propensity for type 2 diabetes are
 CC based on detection of a polymorphism in a calpain encoding nucleic
 CC acid, especially UCSNP-10 of the CAPN10 gene. Methods are also
 CC claimed for: producing calpain 10 polypeptides using calpain
 CC 10-encoding polynucleotides, including the present sequence;
 CC identifying modulators of calpain activity using calpain 10
 CC polypeptides and nucleic acids; and using these modulators to
 CC treat diabetes, in particular through the regulation of an insulin
 CC secretory response or insulin mediated glucose transport.
 XX Sequence 2516 BP; 445 A; 777 C; 837 G; 457 T; 0 other;

alignment_scores:

Quality: 3239.50 Length: 745
 Ratio: 5.267 Gaps: 2
 Percent Similarity: 82.550 Percent Identity: 82.416

alignment_block:

US-09-768-877-2 x AAA27480 ..

Align seg 1/1 to: AAA27480 from: 1 to: 2516

1 MetArgAlaGlyArgGlyAlaThrProAlaArgGluLeuPheArgAspAl 17
 42 ATGCGGG 91
 17 alaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrProL 34
 92 CGCCTTCCCGCGCGGACTCTCGCTCTCTCGGACTTGTCTACGCGCG 141
 34 euAlaGlnPheArgGluAspIleThrTrpArgArgProGlnGluIleCys 50
 142 TGGCCCAAGTTCCGCGAGGACATCAGCTGGAGCGCGCCCGCAGGAGATTGT 191
 51 AlaThrProArgLeuPheProAspAspProArgGluGlyGlnValLysG 67
 192 GCCACACCCCGGCTGTTCCAGATGACCCACCGGAGGCGAGGTGAAGCA 241
 67 nGlyLeuLeuGlyAspCysTrpPheLeuCysAlaCysAlaLeuGlnL 84
 242 GGGGCTGTGGGGGATGCTGGTTCTCTGTGTGCTTGGCGCGCGCTGCAGA 291

84 ySserArgHisLeuLeuAspGlnValIleProProGlyGlnProSerTrp 100
 292 AGAGCAGGCACCTCTCTGGACCAGGTCTTCCTCCGGGACAGCCGAGCTGG 341
 101 AlaAspGlnGluTyrArgGlySerPheThrCysArgIleTrpGlnPheG 117
 342 GCGCACCAGGAGTACCGGGGCTCTTCACCTGTCCGCTTGGCAGTTGG 391
 117 yArgTrpValGluValThrThrAspAspArgLeuProCysLeuAlaGlyA 134
 392 ACGCTGGGTGGAGGTGACACACAGATGACCGCTGCGCTTGCAGGGA 441
 134 rGluCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
 442 GACTCTGTTTCTCCCGCTCCAGAGGAGGATGTGTCTGGCTCCCTTA 491
 151 LeuGluLysValTyrAlaLysValHisGlySerTyrGluHisLeuTrpAl 167
 492 CTGGAAGAGTCTAGCCCAAGGTCCATGGGTCTACGACGACCTGTGGGC 541
 167 aGlyGlnValAlaAspAlaLeuValAspLeuThrGlyGlyLeuAlaGluA 184
 542 CGGGCAGGTGGCGGATGCCCTGTGTGACCTGACCGGCGGCTGGCAGAA 591
 184 rGTrpAsnLeuLysGlyValAlaGlySerGlyGlyGlnGlnAspArgPro 200
 592 GATGAACTGAAAGCGGTAGCAGGAAGCGAGGCGGACGACGACAGGCCA 641
 201 GlyArgTrpGluHisArgThrCysArgGlnLeuLeuHisLeuLysAspG 217
 642 GGCCTCTGGGAGCACAGGACTTGTGGCAGCTGTCCACCTGAAGGACCA 691
 217 nCysLeuSerCysCysValLeuSerProArgAlaGlyAlaArgGluL 234
 692 GTGTCTGATCAGCTGCTGCTCAGCCGCCAGAGCAGGTGCCCGGAGC 741
 234 euGlyGluPheHisAlaPheIleValSerAspLeuArgGluLeuGlnGly 250
 742 TGGGGGAGTTCCATGCTTCATGTCTCGACCTCGGGAGGCTCCAGGCT 791
 251 GlnAlaGlyGlnCysIleLeuLeuLeuArgIleGlnAsnProTrpGlyAr 267
 792 CAGCGGGGCCAGTGCATCTGCTGCTGGGATCCAGNACCCCTGGGGCGG 841
 267 gArgCysTrpGlnGlyLeuTrpArgGluGlyGlyGlyTrpSerGlnV 284
 842 GCGGTCTGGCAGGGGCTCTGGAGAGAGGGGGTGAAGGGTGGAGCCAGG 891
 284 alaPAlaAlaValAlaSerGluLeuLeuSerGlnLeuGlnGluGlyGlu 300
 892 TAGATGACAGGGTAGCATCTGAGCTCTCTGCTCCAGCTCCAGAGGGGAG 941
 301 PheTrpValGluGluGluPheLeuArgGluPheAspGluLeuThrVa 317
 942 TTCTGGGTGGAGGAGGAGGTCTCTCAGGGAGTTTGACGAGCTCACCGT 991
 317 lGlyTyrProValThrGluAlaGlyHisLeuGlnSerLeuTyrThrGluA 334
 992 TGGCTACCCCGGTACGAGGAGGGCGCCACTGCAGAGCCTCTACACAGAGA 1041
 334 rGLeuLysCysHisThrArgAlaLeuProGlyAlaTrpValLysGlyGln 350
 1042 GGCTGCTCGCATACGCGGGCGCTGCTCTGGGGCTGGGTCAAGGGCCAG 1091
 351 SerAlaGlyGlyCysArgAsnAsnSerGlyPheProSerAsnProLysPh 367
 1092 TCAGCAGGAGGCTGCCGGAACAACAGCGGCTTTCACGACCAACCCCAATT 1141
 367 eTrpLeuArgValSerGluProSerGluValTyrIleAlaValLeuGlnA 384
 1142 CTGGCTGGGGTCTCAAAACCCAGGTGAGGTGTACATTGCCGCTCTGCAGA 1191

```

384 rGSeRArgLeuHISAlaAlaSPTRAlaGlyArgAlaAlaLeuVal 400
|||||
1192 GATCCAGCGGCTGACGCGGACTGGCGGCGGCGGCGGCGGCGGCGGCTG 1241
401 GLYASpSerHisThrSerTrpSerProAlaSerIleProGlyLysHisTy 417
|||||
1242 GGTGACAGTCATCTTCTGCGAGCCCGACGACATCCCGGCGACAGCTA 1291
417 rGlnAlaValGlyLeuHISLeuTrpLys 426
|||||
1292 CCAGGCTGTGGTCTGACCTCTGGAAGGGGTGTGACACTTGGACACAC 1341
426 426
1342 TGTTCCTCTCTCTTCATGATGTGGCCACATGATGTTCTTCTCTCT 1391
426 426
1392 GCAGAAAGAGTTCGTGGAAGGCCACTGTCCAGACGCCCGCAGTTGCT 1441
426 426
1442 GGGCCACGGTGGCTTTGTGGCCACAGCTACAGAGAGACTTGGAGGCT 1491
427 427
1492 TGTCTGGGACAGATCTGGCGCCAGGCGCCAGTGAAGCCCGGAGTTGTA 1541
428 GLUTySArgArgValAsnLeuProArgValLeuSerMetProProValAl 444
|||||
1542 GAGAAAGCGCGGCTCATCTGCTAGGGTCTGTCATGATCCCGCGTGGC 1591
444 aGlyThrAlaCysHisAlaTyrAspArgGlyValHisLeuArgCysGlu 461
|||||
1592 TGGCACCCGCTGCATGATGATGACGCGGAGGTCCACCTGGCTTGAGAC 1641
461 euserProGlyTyrTyrLeuAlaValProSerThrPheLeuLysAspAla 477
|||||
1642 TCTCACCGGGCTACTACTGCTGCTCCACGACCTTCTCTGAAGACGCG 1691
478 ProGlyGluPheLeuLeuArgValPheSerThrGlyArgValSerLeu 494
|||||
1692 CCAGGGAGTCTCTCTCGAGTCTTCTTACGGGCGAGTCTCCCTTAG 1740
494 rAlaIleArgAlaValAlaLysAsnThrThrProGlyAlaAlaLeuPro 511
1741 1741
511 IaGlyGluTrpGlyThrValGlnLeuArgGlySerTrpArgValGln 527
1741 1741
528 ThrAlaGlySerArgAsnPheAlaSerTyrProThrAsnProCysPh 544
1741 1741
544 eProPheSerValProGlyGlyProArgCysValArgIleThrL 561
1742 1742
561 euhISGlnHisCysArgProSerAspThrGluPheHisProIleGlyPhe 577
|||||
1773 TGCATTCAGACTGCCGCGCCAGTACACCGAGTTCACCCATCGGCTTC 1822
578 HisIlePheGlnValProGluGlyArgSerGlnAspAlaProPhe 594
|||||
1823 CATATCTTCCAGGTCCAGAGGTGAAGAGGACGACGACGCCACT 1872
594 uLeuLeuGlnGluProLeuLeuSerCysValProHisArgTyrAlaGln 611
|||||
1873 GCTGCTCCAGAGCGGCTGCTGAGACTGCTGCCACATGCTACAGCCCG 1922
611 LuValSerArgLeuCysLeuLeuProAlaGlyThrTyrLysValValPro 627

```

```

|||||
1923 AGCTGAGCGCGGCTCTGCTCTGTCGAGGACCTACAGAGTTGCGCC 1972
628 SerThrTyrLeuProAspThrGluGlyAlaPheThrValThrIleAla 644
|||||
1973 TCCACCTACCTCCGCGACACAGAGGGGCGCTTCACAGTACATCCGAC 2022
644 rArgIleAspArgProSerIleHisSerGlnGluMetLeuGlyGlnPhe 661
|||||
2023 CAGGATTGACAGGCGCATCTTCACACCGAGGAGATGCTGGCGGCTCC 2072
661 euGlnGluValSerValMetAlaValMetLysThr 672
|||||
2073 TCCAGAGGCTCTCCGTCATGCGCAGTGTGAAACC 2107

seq_name: /SIDSL/ycgdata/geneseq/geneseqn-emb1/NA2000.DAT.AAA27479

seq_documentation_block:
ID AAA27479 standard; cDNA: 2204 BP.
XX
AC AAA27479;
XX
DT 15-AUG-2000 (first entry)
XX
DE Human calpain 10d cDNA.
XX
KW NIDDM1: non-Insulin-dependent diabetes mellitus; CAPN10 gene;
KW calpain 10; calpain 10d; diapain-1; cysteine protease;
KW chromosome 2; human; diagnosis; therapy; ss.
XX
OS Homo sapiens.
XX
FH Key 42..1583
FT CDS /*tag= a
FT
PN W0200023603-A2.
XX
PD 27-Apr-2000.
XX
PE 21-OCT-1999; 99WO-US24890.
XX
PR 21-OCT-1998; 98US-0105052.
PR 13-MAY-1999; 99US-0134175.
XX
PA (ARCH-) ARCH DEV CORP.
XX
PI Polonsky KS, Horikawa Y, Oda N, Sreenan S, Zhou Y, Otani K;
PI Hanis CL, Bell GI;
XX
DR WPI: 2000-339702/29.
XX
P-PSDB: AAY79570.
XX
PT Method for screening for type 2 diabetes mellitus comprises detecting a
PT polymorphism in a calpain encoding nucleic acid segment or a
PT protease-encoding nucleic acid segment -
XX
PS Claim 65; Page 227; 257pp; English.
XX

The present sequence is that of cDNA corresponding to a minor
transcript of the human calpain 10 gene, CAPN10 (see AA227475).
The sequence contains an open reading frame that encodes a protein
of 513 amino acids (see AAY79570), designated calpain 10d. The
CAPN10 gene consists of 15 exons spanning 32 kb. A complex pattern
of alternative splicing generates proteins of 672, 544, 517, 513,
444, 274, 139 and 138 amino acids (see AAY79567-74), designated
calpain 10a to 10h, respectively. Calpain 10d is encoded by exons
1-7, 9 and 11-13 of the gene. The invention concerns the
identification of genes responsible for type 2 diabetes for use in
diagnostic and therapeutic applications. A G-to-A polymorphism in
intron 3 of the CAPN10 gene in the NIDDM1 region of chromosome 2
(CC (UCSNP-43) shows evidence for linkage to type 2 diabetes. Claimed
methods for screening for a propensity for type 2 diabetes are

```

CC based on detection of a polymorphism in a calpain encoding nucleic
 CC acid, especially UCSNP-10 of the CAPN10 gene. Methods are also
 CC claimed for: producing calpain 10 polypeptides using calpain
 CC 10-encoding polynucleotides, including the present sequence;
 CC identifying modulators of calpain activity using calpain 10
 CC polypeptides and nucleic acids; and using these modulators to
 CC treat diabetes, in particular through the regulation of an insulin
 CC secretory response or insulin mediated glucose transport.
 XX
 SQ Sequence 2204 BP; 390 A; 672 C; 752 G; 390 T; 0 Other;

alignment_scores:
 Quality: 3090.50 Length: 672
 Ratio: 5.292 Gaps: 1
 Percent Similarity: 86.905 Percent Identity: 86.905

alignment_block:

US-09-768-877-2 x AAA27479 ..

Align seg 1/1 to: AAA27479 from: 1 to: 2204

1 MetArgAlaGlyArgGlyAlaThrProAlaArgGluLeuPheArgAspAl 17
 42 ATCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGG 91
 17 aAlaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrProL 34
 92 CGCTTCCCGCGCGCGGACTCTCTCGCTCTCTCGGACTTGTCTACCGCG 141
 34 euAlaGlnPheArgGluAspIleThrTrpArgArgProGlnGluIleCys 50
 142 TGGCCAGTTCCTCCGAGGACATCAGCTGGAGGGCGGGCGGGCGGG 191
 51 AlaThrProArgLeuPheProAspAspProArgGluGluValIysG 67
 192 GCCACACCGCGGTGTTCAGATACCCAGGAGGGCGGGCGGGCGGG 241
 67 nGlyLeuLeuGlyAspCysTrpPheLeuCysAlaCysAlaAlaLeuGln 84
 242 GGGGCTGTGGGGATTGTGGTTCCTGTCTGCTCGCGCGGGCGGG 291
 84 ySerArgHisLeuLeuAspGlnValIleProProGlyGlnProSerTrp 100
 292 AGAGCAGGACCTCTCTGGACAGGATCATTCCTCGGGCGGGCGGG 341
 101 AlaAspGlnGluTrpArgGlySerPheThrCysArgIleTrpGlnPheG 117
 342 GCCGACCGAGGTACCGGGGCTCTTCACTGTGCGATTGGCAGTTGG 391
 117 yArgTrpValGluValThrTrpAspAspArgLeuProCysLeuAlaGly 134
 392 ACCTGGGTGGAGGTGACACACAGATGACCGCTGCGCTGCGGTGG 441
 134 rGlyCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
 442 GACTCTGTCTCTCCCGCTCCAGAGGGAGGATGTGTCTGGCTCCCTTA 491
 151 LeuGluLysValTrpAlaLysValHisGlySerTrpGluHisLeuTrpAl 167
 492 CTGGAAGAGGTCTACCCAAAGTCTCATGGGTCTTACGAGCCTGTGG 541
 167 aGlyGlnValAlaAspAlaLeuValAspLeuThrGlyGlyLeuAlaGlu 184
 542 CGGGCAGGTGGCGGATGCTGTGGTGTGACCGCGGGCGGGCGGG 591
 184 rGTrpAsnLeuLysGlyValAlaGlySerGlyGlnGlnAspArgPro 200
 592 GATGGAACCTGAAGGGGCTAGCAGGAAGCGGAGCGGAGGAGGAGG 641
 201 GlyArgTrpGluHisArgThrCysArgGlnLeuLeuHisLeuLysAsp 217
 642 GCGCGCTGGGAGCACAGGACTTGTGGGAGCTGTCTCCACCTGAAGG 691

217 nCysLeuIleSerCysCysValLeuSerProArgAlaGlyAlaArgGlu 234
 692 GTCTCTGATCAGCTGTGCTGTCTGCTCAGCCAGAGCAGAGTGTCCCGGAGC 741
 234 euGlyGluPheHisAlaPheIleValSerAspLeuArgGluLeuGlnGly 250
 742 TGGGGAGGTTCATGCTTCTTCGAGACCTTGGGGAGCTTCCAGGGT 791
 251 GlnAlaGlyGlnCysIleLeuLeuLeuArgIleGlnAsnProTrpGlyAr 267
 792 CAGCGGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 841
 267 gArgCysTrpGlnGlyLeuTrpArgGluGlyGlyGlyGlyGlyGly 284
 842 CGGGTGTGCGAGGGGCTCTGAGAGAGGGGGGTGAGGGTGGAGCCAGG 891
 284 aAlaAlaAlaValAlaSerGluLeuLeuSerGlnLeuGlnGluGlyGlu 300
 892 TAGATCGACGGGTAGCATCTGAGCTCTGTCCAGCTTCCAGGAGGGGAG 941
 301 PheTrpValGluGluGluPheLeuArgGluPheAspGluLeuThrVa 317
 942 TTTCTGGGTGGAGGAGGAGGAGTTCCTCAGGGAGTTTGACGAGCTCACCGT 991
 317 lGlyTrpProValThrGluAlaGlyHisLeuGlnSerLeuTrpThrGlu 334
 992 TGCTTACCGGGTCTCAGGAGGGCGGGCGGGCGGGCGGGCGGGCGGG 1041
 334 rGluLeuCysHisThrArgAlaLeuProGlyAlaTrpValLysGlyGln 350
 1042 GGCTGTCTGTCCATACGGGGCGCTGCTGCGGGCGGGCGGGCGGGCGGG 1091
 351 SerAlaGlyGlyCysArgAsnAsnSerGlyPheProSerAsnProLysPh 367
 1092 TCAGCAGGAGGCTGCGGGAACACAGCGGCTTTCCAGCAACCCCAATT 1141
 367 eTrpLeuArgValSerGluProSerGluValTrpIleAlaValLeuGln 384
 1142 CTGGCTTGGGGTCTCAGAACCGAGTGGGTGTACATTGCGCTCTGCAGA 1191
 384 rGSerArgLeuHisAlaAlaAspTrpAlaGlyArgAlaArgAlaLeuVal 400
 1192 GATCCAGGCTGCAGCGGGCGGCTGCGGCGGGCGGGCGGGCGGGCGGG 1241
 401 GlyAspSerHisThrSerTrpSerProAlaSerIleProGlyLysHisTy 417
 1242 GGTGACGTATCTTCTGTGGAGCCAGCGAGCATCCCGGGCAAGCACTA 1291
 417 rGlnAlaValGlyLeuHisLeuTrpLysValGluLysArgArgValAsnL 434
 1292 CCAGGCTGTGGGTCTGCACTCTTGAAGGTAGAGAGCGGGCGGGGTCAATC 1341
 434 euProArgValLeuSerMetProValAlaGlyThrAlaCysHisAla 450
 1342 TGCTTAGGGTCTGTCTTCCATGCCCCCGTGGTGGCAGCGGTGCCATGCA 1391
 451 TyrAspArgGluValHisLeuArgCysGluLeuSerProGlyTyTrLe 467
 1392 TACGACCGGGAGGTCCACCTGCTGTGAGCTCTCAGCGGGCTACTACCT 1441
 467 uAlaValProSerThrPheLeuLysAspAlaProGlyGluPheLeuLeu 484
 1442 GGCTGTCCCCAGCACTTCTTGAAGAGCGCGGAGGGAGTTCCTGCTCC 1491
 484 rGValPheSerThrGlyArgValSerLeuSerAlaIleArgAlaValAla 500
 1492 GAGTCTTCTTACCGGGCGGAGTCTCCCTTAG..... 1522
 501 LysAsnThrThrProGlyAlaAlaLeuProAlaGlyGluTrpGlyThrVa 517
 1522 1522

134 rgLeuCyspheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
509 GACTCTGCTTTTCCGGGTGCCAGAGAGAGGATGTGTCTGGCTTCCCTTA 558
151 LeuGluLysValTyrAlaLysValHisGlySerTyrGluHisLeuTrpAl 167
559 CTGGAAAAGCGCTATGCTTAAGTCCATGGATCGATATGAGCACCTGTGGC 608
167 acGlyGlnValAlaAspAlaLeuValAspLeuThrGlyGlyLeuAlaGluA 184
609 AGGCCAAGTGCAGATGCTTAGTGGATCTCACTGGAAGCCTGCAGAA 658
184 rgTrpAsnLeuLysGlyValAlaGlySerGlyGlyGlnGlnAspArgPro 200
659 GGTGGAGCTTCAAGGATGTAAACAAAGCCAGCGCCACAGCAGCAGACC 708
201 GlyArgTrpGluHisArgThrCysArgGlnLeuLeuHisLeuLysAspGl 217
709 AGTGTGGGAGCACAGACTGTGCGCAGCTACTCCACTGAAGGACCG 755
217 nCysLeuIleSerCysCysValLeuSerProArgAlaGlyAlaArgGluL 234
759 GTGTCTAATCAGCTGCTCTGTGCTTAGCCCCAGACGAGGTGCCAGGAAC 808
234 euGlyGluPheHisAlaPheIleValSerAspLeuArgGluLeuGlnGly 250
809 TCGGAGAGTTCCATGCTCTCATCATCTCAGATCTGCAGGAGCTCAGGAGT 858
251 GlnAlaGlyGlnCysIleLeuLeuLeuArgIleGlnAsnProTrpGlyArg 267
859 CAGACTGGCCAGGATATCTCTCTGCTCGGATTCACAACCCCTGGGGCCG 908
267 gArgCysTrpGlnGlyLeuTrpArgGluGlyGlyGluGlyTrpSerGlnV 284
909 CGGTGTGTGGCAGGCGCTTCGGAGAAAGAGAGGTGAAGGTGGAACCCAGG 958
284 aAlaAspAlaValAlaSerGluLeuLeuSerGlnLeuGlnGlyGlyGlu 300
959 TAGAGCCAGTCAAGGAGTCTGAGCTGTGCGCCCACTCCAGGAAGGAGAG 1008
301 PheTrpValGluGluGluPheLeuArgGluPheAspGluLeuThrVal 317
1009 TCTGTGGTTCGAGGAAGAGGAGTTCTCTCAGGAGGTTGTATGAGTCCACAT 1058
317 lGlyTyrProValThrGluAlaGlyHisLeuGlnSerLeuTyrThrGluA 334
1059 CGGCTACCCAGTACAGAGCGCCGCCACCTACAGAGTCTCCACACAGAGA 1108
334 rgLeuLeuCysHisThrArgAlaLeuProGlyAlaTrpValLysGlyGln 350
1109 GGTGTGTGTCATACCGGACACTGCTGTGGTGCCTGGGTGACAGGGCAG 1158
351 SerAlaGlyGlyCysArgAsnAsnSerGlyPheProSerAsnProLysPh 367
1159 TCAGCAGAGGCTGCCGGAACAACAGTGTCTTCCCTGCAACCCCAAGTT 1208
367 eTrpLeuArgValSerGluProSerGluValTyrIleAlaValLeuGlnA 384
1209 CTGCTTACGCTCTTGGAAACCCAGCAGGAGTGTGTGGCTGTCTTCAGA 1258
384 rgSerArgLeuHisAlaAlaAspTrpAlaGlyArgAlaArgAlaLeuVal 400
1259 GACCCCGGAGCGCTTA.....GTGGGCCAGACTCGGGCAGCTGGCG 1299
401 GlyAspSerHisThrSerTrpSerProAlaSerIleProGlyLysHisTy 417
1300 GGT.....GCCAGTCTCGACCGGTGAACCTCCCGAGGCAAGACTA 1340
417 rGlnAlaValGlyLeuHisLeuTrpLysValGluLysArgArgValAsnL 434
1341 CCAGCTGTGGGCTGCACATCTGGAAGGTGAGAAGACCGAAGACTCAGCC 1390
434 euProArgValLeuSerMetProProValAlaGlyThrAlaCysHisAla 450

1391	TGCCCAGAGTCTGTCTGCACCCCTGGCTGGCACTGATGCATGCAC	1440
451	TyrAspArgGluValHisLeuArgCysGluLeuSerProGlyTyrrTrile	467
1441	TATGATCGTGAGATCCACTTCGGTTGTGAGCTCTCACAGGCTACTACCT	1490
467	uaLaValProSerThrPheLeuLysAspAlaProGlyGluPheLeuLeuA	484
1491	GGCCGTCCCTACACACTTTTTGAAGGATGTGCCAGGCGACTTCTGTCTCA	1540
484	rgValPheSerThrGlyArgValSerLeuSerAlaIleArGAlaValAla	500
1541	GAGTCTTCTACTGGGAATCTCCCTCATGCTGGCTGAGCTGGCCACC	1590
501	LysASNThrThrProGlyYalaAlaLeuProAlaGlyGluTrpGlyThrVa	517
1591	AAGGTTGCATCGCTTGGAAACGCCCTGCCCTCAGCCAGTGGGNAGCTGT	1640
517	IgLnLeuArgGlySerTrpArgValGlyGlnThrAlaGlyCylserArgA	534
1641	GCAGTTGTGAGGGCTCTGGAGAGCTGGCCACACAGCTGGGGCAGCAGGA	1690
534	snPheAlaSerTyrrProthrAsnProCysPheProPheSerValProGlu	550
1691	ACTTTGCCCTTCTAACCCCTGCAATCCCTGCCCTCCCTTCTCTGTCCCTGAG	1740
551	GlyProGlyProArgCysValArgIleThrLeuHisGlnHisCysArgPr	567
1741	GGTGTGCCCCCGCTACATCCGTATCACCTCAGACACACACTGCCGGCT	1790
567	oSerAspThrGluPheHisProIleGlyPheHisIlePheGlnValProg	584
1791	CAGTGACAGCCAGCTGCACCCCATTTGGTTTCCATGTCTTTCAGGTTCCAG	1840
584	lucIGlyArgSerGlnAspAlaProPheLeuLeuGlnGluProLeu	600
1841	CAGACGGTGAGAACCAAGACGCGTGTTCCTGTCTGTCTCCAGGAGCCACTG	1890
601	LeuSerCysValProHisArgTyrrAlaGlnGluValSerArgLeuCysLe	617
1891	CTAAGCTGTGTACCACATCGCTACGCCCAGGAGTGAGCGCCTCTCGCCT	1940
617	uLeuProAlaGlyThrTyrrLysValValProSerThrTyrrLeuProAspt	634
1941	CCCTTCTGTGGGAACTACAGGATGTTCCTCCACCTACCTGCCAGATA	1990
634	hrGlucyAlaPheThrValThrIleAlaThrargIleAsparqProSer	650
1991	CAGAGGGTACCTTCACGGTAACCATACACCAAGAATCGATAGGCGAGTCC	2040
651	IleHisSerGlnGluMetLeuGlyGlnPheLeuGlnGluValSerValMe	667
2041	ATCCACACCCAGGAGATGCTGGCGCAGCTGCTCCAGGAGGTCTCCCTTTAT	2090
667	talaValMetLys	671
2091	GGCAGTGATGAAA	2103

seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:AAS97171

seq_documentation_block:

ID AAS97171 standard; cDNA; 1542 BP.

XX
AC

XX
XX
T/T/CCVV

DT 26-FEB-2002 (first entry)

XX
DE Human cysteine protease partial DNA sequence #9.

Human: cytochrome P-450 partial DNA sequence #9.
XX
KW Human: protease; PCR primer; cytostatic; immunomodulator; cardiant;
KW vasotropic; antimigraine; analgesic; endocrine; nootropic; tranquiliser
KW hypertensive; hypotensive; neuroleptic; neuroprotective; anabolic;
KW

351 SerAlaGlyGlyCysArgAsnAsnSerGlyPheProSerAsnProLysPh 367
1051 TCACGAGGCTGCGGGAACAACAGCGGCTTTCCACGACACCCCAATT 1100
367 eTrpLeuArgValSerGluProSerGluValTyrIleAlaValLeuGlnA 384
1101 CTGGCTCGGGTCTCAGAACCGAGTGGTGTACATGTCGCTCGTCGAGA 1150
384 rGSerArgLeuHisAlaAlaAsTrpAlaGlyArgAlaArgAlaLeuVal 400
1151 GATCCAGGCTGCACGGCGGCGACTGGCAGCGCGGCGCGGCGACTGGTG 1200
401 GlyAspSerHisThrSerTrpSerProAlaSerIleProGlyLysHisTy 417
1201 GGTGACAGTCATCTTGTGTGAGCCCGCAGCGAGCATCCGGGCAAGCACTA 1250
417 rGlnAlaValGlyLeuHisLeuTrpLysValGluLysArgArgValAsnL 434
1251 CCAGGCTGTGGGCTGCACCTCTGGAGGAGTAGAGAGCGGGGTCAATC 1300
434 euProArgValLeuSerMetProProValAlaGlyThrAlaCysHisAla 450
1301 TGCTAGGGTCCGTGTCATGCCCGCCCGTGGCTGGCAGCGGCGCATGCA 1350
451 TyrAspArgGluValHisLeuArgCysGluLeuSerProGlyTyrTyrLe 467
1351 TACGACCGGAGGTCCACCTGCTGTGTGAGCTCTCACCGGGCTACTACCT 1400
467 uAlaValProSerThrPheLeuLysAspAlaProGlyGluPheLeuLeuA 484
1401 GGTGTCTCCCGACACCTTCTGAAGGAGCGCCAGGGGAGTTCCTGCTCC 1450
484 rGValPheSerThrGlyArgValSerLeuSerAlaIleArgAlaValAla 500
1451 GAGTCTCTCTACCGGGCGAGTCTCCCTTAG 1481
501 LysAsnThrThrProGlyAlaAlaLeuProAlaGlyGluTrpGlyThrVa 517
1481 1481
517 lGlnLeuArgGlySerTrpArgValGlyGlnThrAlaGlyGlySerArgA 534
1481 1481
534 snPheAlaSerTyrProThrAsnProCysPheProPheSerValProGlu 550
1481 1481
551 GlyProGlyProArgCysValArgIleThrLeuHisGlnHisCysArgPr 567
1481 1481
567 oSerAspThrGluPheHisProIleGlyPheHisIlePheGlnValProG 584
1482GTCCCGAG 1488
584 luGlyGlyArgSerGlnAspAlaProProLeuLeuGlnGluProLeu 600
1489 AGGGTGAAGGAGCAGGAGCAGCAGCCCGGCTGCTGCTGAGGAGCGGCTG 1538
601 Leu 601
1539 CTG 1541
seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AAD13883
seq_documentation_block:
ID AAD13883 standard; cDNA: 2180 BP.
XX AAD13883;
AC
XX
DT 29-NOV-2001 (first entry)
XX

DE XX Human calpain-like protease, 18036 cDNA.
KW Calpain-like protease 18036; spleen disorder; splenomegaly; lung;
KW adult respiratory distress syndrome; colon; liver; jaundice; brain;
KW idiopathic inflammatory bowel disease; cerebrovascular disease;
KW acute meningitis; Alzheimer's disease; T-cell; transplant rejection;
KW systemic lupus erythematosus; skin; seborrhoeic keratosis; leukaemia;
KW haematopoietic stem cell; heart; myocardial infarction; atherosclerosis;
KW kidney; acute proliferative glomerulonephritis; urolithiasis; apoptosis;
KW ischaemia; neurodegenerative disease; demyelinating disease; injury; EAE;
KW experimental allergic encephalomyelitis; multiple sclerosis; spinal cord;
KW LGMD2A muscular dystrophy; proliferative disorder; cancer; gene therapy;
human; chromosome 2; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 183..1736
FT /*tag= a
FT /product= "Calpain-like protease, 18036"
FT /note= "This region is claimed as SEQ ID NO: 3 in
FT claim 1."
XX
PN WO200164919-A2.
XX
XX 07-SEP-2001.
XX
XX 26-FEB-2001; 2001WO-US06110.
XX
XX 28-FEB-2000; 2000US-0185333.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX Kapeller-Libermann R;
XX
XX WPI; 2001-565507/63.
XX P-PSDB; NAE09756.
XX
XX New calpain-like protease polypeptides and polynucleotides for
XX diagnosing, treating seven-transmembrane protein/receptor-related
XX disorders and to identify modulators of therapeutic use
XX
XX Claim 1; Fig 1; 114pp; English.
XX
XX The invention relates to polypeptide and polynucleotide of calpain-like
XX protease, 18036. The polypeptide and polynucleotide of the invention are
XX useful for treatment and diagnosis of calpain-like protease-mediated
XX disorders. These disorders include disorders of spleen (e.g.
XX splenomegaly), lung (e.g. adult respiratory distress syndrome), colon
XX (e.g. idiopathic inflammatory bowel disease), liver (e.g. jaundice),
XX brain (e.g. traumatic brain injury, cerebral oedema, cerebrovascular
XX disease, acute meningitis, Alzheimer's disease), T-cells (e.g. transplant
XX rejection, systemic lupus erythematosus), skin (e.g. seborrhoeic
XX keratosis), haematopoietic stem cells (e.g. leukaemias), heart (e.g.
XX myocardial infarction, atherosclerosis), kidney (e.g. acute proliferative
XX glomerulonephritis, urolithiasis), thymus, breast, testis, uterus,
XX epididymis, endometrium, prostate, thyroid, skeletal muscle, pancreas,
XX small intestine, disorders related to reduced platelet number, bone,
XX ovary, pain and infectious disorders. The molecules of the invention are
XX also useful for diagnosing and treating disorders associated with
XX perturbed cellular growth and differentiation, exercise-induced injury
XX and repair, apoptosis, ischaemia, neurodegenerative diseases,
XX demyelinating diseases including experimental allergic encephalomyelitis
XX (EAE) and multiple sclerosis. LGMD2A muscular dystrophy, spinal cord
XX injury, proliferative and differentiative disorders e.g. cancer and
XX renal cell death associated with diverse toxicants. The present sequence
XX is human calpain-like protease, 18036 cDNA. The 18036 gene is located on
XX chromosome 2 between D2S140 and D2S2338.
XX
XX Sequence 2180 BP; 395 A; 651 C; 764 G; 370 T; 0 other;
XX
XX alignment_scores:

Quality:	2719.50	Length:	672
Ratio:	5.260	Gaps:	1
Percent Similarity:	76.935	Percent Identity:	76.935

alignment_block:
US-09-768-877-2 x AAD13883

Align seg 1/1 to: AAD13883 from: 1 to: 2180

1 MetAValaIaIyrgIyValaIthrProIaArgIuIeuhAeAaPaI 17
 183 ATGCGGGCGGGCGGGCGGAGCCCGGCGAGGGAGGTCTTCGGGAGCG 232
 17 aAlaPheProIaAlaAspSerSerLeuPheCysAspLeuSerThrProI 34
 233 CGCCTTCGCCGCGGAGACTCTCGCTCTTCGTGGACTTGTCAAGCGCG 282
 34 euAaIaGlnPheAaArgGluAspIleThrTrpArgArgProGlnGluIleCys 50
 283 TGGCCCAAGTCCCGCGAGGACATACGCGGAGGGGCCCCAGGAGATTGTT 332
 51 AlaThrProArgLeuPheProAspAspArgGlnGluValIleValIlysgI 67
 333 GCCACAGCCCGCGCTTTTCCAGATGACCCAGGMAAGGCAAGGTGAAGCA 382
 67 nGlyLeuLeuGlyAspCysTrpPheLeuCysAlaCysAlaIleuGlnI 84
 383 GGGGCTGCTGGGGGATTGGCTGCTCTGTGCTGCTGGCGCGGCTGCAGA 432
 84 yASerTrpHisLeuAaAspGlnValIleProProGlyGlnProSerTrp 100
 433 AGAGCAGGACCTCCCTGGAGCAGGTATTCCTCCGGAGACGCGACCTGG 482
 101 AlaAspGlnGluTrpArgIlySerPheThrCysArgIleTrpGlnPheG 117
 483 GCGGACAGGAGTACCGGGGGCTCTTACCCTGCGCATTTGGCAGATTGG 532
 117 yArgTrpValGluValThrAspAspArgLeuProCysLeuAlaGlu 134
 533 AGCCTGGTGGAGGTGACCAAGATGACCGCTGGCGCTTGGAGGGA 582
 134 rGLeuCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
 583 GACTCTGTTCCTCCGCTCCGAGGAGGAGATGTCTTGGGCTCCCTTA 632
 151 LeuGluLysValTyrAlaIysValHisGlySerTyrGlnHisLeuTrpAl 167
 633 CTGGAAAGAGCTTACGCCCAAGGTCCATGGGTCTTACAGACCACTGTGGC 682
 167 aGlyIleValAlaAspAlaLeuValAspLeuThrGlyIleuAlaGluA 184
 683 CGGGCAGGTGGCGGATGCCCTGTGGACTGCACGGCGGCGCTGGGANA 732
 184 rGTAspAspLeuLysGlyValAlaGlySerGlyGlyGlnAspArgPro 200
 733 GATGGACCTGAAAGGCGGTAGAGGAAAGGGAGGCGCAGCAGCAGGCA 782
 201 GlyArgTrpGlnHisArgThrCysArgGlnLeuLeuHisIleuLysAspI 217
 783 GGGCCCTGGAGCAACAGACTTGTGGCAAGCTGTCTCCACCTTAAGACCA 832
 217 nCysLeuIleSerCysCysValLeuSerProArgAlaGluAaArgGlu 234
 833 GTGTCTGATCAGCTCGTGGTCTCAGGCCAGACAGTGGCCGGGAGC 882
 234 euGlyGluPheHisAlaPheIleValSerAspLeuArgGluLeuGlnGly 250
 883 TGGGGGAGTTCATGCCCTTCATTGTCTGGACTCGGAGAGCTCCAGGT 932
 251 GlnAlaGlyGlnCysIleLeuLeuLeuArgIleGlnAsnProTrpGlyArg 267
 933 CAGCGGGCGCAGTGCATCTGTGTGGAGATCAGAAACCCCTGGGGCGG 982

[illegible]

```
1461 .....|
584 luGlyGlyArgSerGlnAspAlaProProLeuLeuGlnGlnProLeu 600
1468 AGGGTGAAGAGCCAGGACGACCCCTCTGCTGAGGAGCGCTG 1517
601 LeuSerCysValProHisArgTyrAlaGlnGlnValSerArgLeuCysLe 617
1518 CTCAGCTGGTCCACATCGCTACGCCAGGAGGTGAGCGGCTCTGCT 1567
617 uLeuProAlaGlyThrTyrLysValValProSerThrThrLeuProAsp 634
1568 CTGCTGCTGGGGCACCCTACAGGTGTGCTCCCTCCACCTACCTGCGG 1617
634 hrGluGlyAlaPheThrValThrIleAlaThrArgIleAspArgProSer 650
1618 CAGAGGGGCTTCACGTGACCATCGACACGAGGATGACAGGCAATCC 1667
651 IleHisSerGlnGluMetLeuGlyGlnPheLeuGlnGlnValSerValMe 667
1668 ATTCACAGCCAGGAGATGCTGGCCAGTTCCTCCCAAGAGGTCTCGTCAT 1717
667 tAlaValMetLysThr 672
1718 GCCAGTGATGAATCC 1733
```

```
seq_name: /SIDS1/qcgdata/geneseq/geneseq-emb1/NA2000.DAT:AAA27478
seq_documentation_block:
ID AAA27478 standard; cDNA; 2001 BP.
XX
AC AAA27478;
```

```
15-AUG-2000 (first entry)
Human calpain 10c cDNA.
XX
DE
XX
XX
XX NIDDM1; non-insulin-dependent diabetes mellitus; CAPN10 gene;
KW calpain 10; calpain 10c; diapaen-1; cysteine protease;
KW chromosome 2; human; diagnosis; therapy; ss.
XX
OS Homo sapiens.
```

```
Key Location/Qualifiers
CDS 42..1595
FT /*tag= a
XX
XX WO200023603-A2.
XX
XX 27-APR-2000.
XX
XX 21-OCT-1999; 99WO-US24890.
XX
XX 21-OCT-1998; 98US-0105052.
XX
XX 13-MAY-1999; 99US-0134175.
XX
XX (ARCH-) ARCH DEV CORP.
XX
XX Polonsky KS, Horikawa Y, Oda N, Sreenan S, Zhou Y, Otani K;
PI Hanis CL, Bell GI;
XX
XX WPI; 2000-339702/29.
XX
XX P-PSDB; AA79569.
```

```
Method for screening for type 2 diabetes mellitus comprises detecting a
PT polymorphism in a calpain encoding nucleic acid segment or a
PT protease-encoding nucleic acid segment -
XX
XX Claim 65; Page 224-225; 257pp; English.
XX
XX The present sequence is that of cDNA corresponding to a
CC transcript of the human calpain 10 gene, CAPN10 (see AA27475).
```

```
CC The sequence contains an open reading frame that encodes a protein
CC of 517 amino acids (see AA79569), designated calpain 10c. The
CC CAPN10 gene consists of 15 exons spanning 32 kb. A complex pattern
CC of alternative splicing generates proteins of 672, 544, 517, 513,
CC 444, 274, 139 and 138 amino acids (see AA79567-74), designated
CC calpain 10a to 10h, respectively. Calpain 10c is encoded by exons
CC 1-7 and 11-13 of the gene. It is readily detectable in many
CC tissues, including skeletal muscle and islets. The invention
CC concerns the identification of genes responsible for type 2 diabetes
CC for use in diagnostic and therapeutic applications. A G-to-A
CC polymorphism in intron 3 of the CAPN10 gene in the NIDDM1 region of
CC chromosome 2 (UCSNP-43) shows evidence for linkage to type 2
CC diabetes. Claimed methods for screening for a propensity for type
CC 2 diabetes are based on detection of a polymorphism in a calpain
CC encoding nucleic acid, especially UCSNP-10 of the CAPN10 gene.
CC Methods are also claimed for: producing calpain 10 polypeptides
CC using calpain 10-encoding polynucleotides, including the present
CC sequence; identifying modulators of calpain activity using calpain
CC 10 polypeptides and nucleic acids; and using these modulators to
CC treat diabetes, in particular through the regulation of an insulin
CC secretory response or insulin mediated glucose transport.
XX
SQ Sequence 2001 BP; 361 A; 602 C; 690 G; 348 T; 0 other;
```

```
alignment_scores:
Quality: 2715.50 Length: 672
Ratio: 5.252 Gaps: 1
Percent Similarity: 76.935 Percent Identity: 76.786
alignment_block:
US-09-768-877-2 x AAA27478
Align seg 1/1 to: AAA27478 from: 1 to: 2001
```

```
1 MetArgAlaGlyArgGlyAlaThrProAlaArgGluLeuPheArgAspAl 17
42 ATGCGGGCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 91
17 aAlaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrPro 34
92 GCGCTTCCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 141
34 euAlaGlnPheArgGluAspIleThrTrpArgArgProGlnGluIleCys 50
142 TGGCCCGAGTTCGCGGAGGACATCATCGTGGAGGCGGCGGCGGCGGCG 191
51 AlaThrProArgLeuPheProAspAspProArgGluGlyGlnValLysG 67
192 GCCACACCGCGGCTGTTCCAGATGACCCAGCGGAGGCGGCGGCGGCGG 241
67 nGlyLeuLeuGlyAspCysTrpPheLeuCysAlaCysAlaAlaLeuGln 84
242 GGGGCTGCTGGGGGATTCGCTGGTTCCTGTGCTGCGCGCGCGCTGCAGA 291
84 ysSerArgHisLeuLeuAspGlnValIleProProGlyGlnProSerTrp 100
292 AGACGAGCACCCTCTCGGACCGGTCATTCCTCGGAGCAGCGGAGCTGG 341
101 AlaAspGlnGluTyrArgGlySerPheThrCysArgIleTrpGlnPheG 117
342 GCGGACGAGGAGTACCGGGGCTCTTCACCTGTCCGATTTGGCAGTTGG 391
117 YArgTrpValGluValThrThrAspAspArgLeuProCysLeuAlaGly 134
392 ACGCTGGGTGGGTGAGTGACACAGATGACCGCTGCGCTTGCAGGGA 441
134 rgLeuCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
442 GACTCTGTTTCTCCCGCTGCCAGAGGAGGATGTGTTCTGCTCCCTTA 491
151 LeuGluLysValTyrAlaLysValHisGlySerTyrGluHisLeuTrpAl 167
```

```

492 CTGAAAAGCTCTACGCCAAGCTCATGGCTCTACGACACCTCTGGGC 541
167 aGlyGlnValAlaAspAlaLeuValAspLeuThrGlyLeuAlaGln 184
      |||
542 CGGCGAGGTGGCGGATGCCCTGGTGACCTGACCGCGCGCTGGAGAA 591
184 rGTPAspLeuLysGlyValAlaGlySerGlyGlyGlnAspArgPro 200
      |||
592 GATGAACCTGAAGGCGCTACAGAACGAGACGACAGACAGAGCCA 641
201 GLyATrTPGlnHisArgThrCysArgGlnLeuLeuHisLeuLysAspG 217
      |||
642 GCGCGCTGGAGCACAGACACTTGTGGCAGCTGCTCCACTGAAGACCA 691
217 nCysLeuLeuSerCysCysValLeuSerProArgAlaGlyAlaArgGln 234
692 GTGTCTGATCAGTGTCTGCTCAGCCCAAGACAGGTGCCCGGAGC 741
234 euGlyGlnPheHisAlaPheIleValSerAspLeuArgGlnLeuGln 250
      |||
742 TGGGGAGATTCCATGCCCTCATTTCTCGGACCTGCGGAGCTCAGGGT 791
251 GlnAlaGlyGlnCysIleLeuLeuLeuArgIleGlnAsnProTPGlyAr 267
      |||
792 CAGCGCGGCCAGTGCATCTGCTGTGGATCCAGAACCCCTGGGGCCG 841
267 gATgCysTrpGlnGlyLeuTrpArgGlnGlyGlyGlyTrpSerGln 284
842 GCGGTGCTGGCAGGGGCTGGAGAGAGGGGGTGAAGGTGGAGCCAGG 891
284 alAspAlaAlaValAlaSerGluLeuLeuSerGlnLeuGlnGlyGln 300
892 TAGATGACGGGTGACATCTGAGCTCTGCCCTGCCAGCTCCAGAGGGGAG 941
301 PheTrpValGlnGlnGlnGlnPheLeuArgGlnPheAspGluLeuTrp 317
942 TTCTGGGTGGAGAGAGAGATTCTCAGGAGATTGACGAGCTCACCGT 991
317 lGlyTrpProValThrGlnAlaGlnHisLeuGlnSerLeuTrpThrGln 334
992 TGGCTACCGGTGACGAGGCGGCGCACAGCTGACAGCTTACACAGAGA 1041
334 rGlyLeuLeuCysHisThrArgAlaLeuProGlyAlaTrpValLysGln 350
1042 GGTGTGCTCTCCATACCGGGCGCTGCGGTGGCTGCAAGGGCCAG 1091
351 SerAlaGlyLysArgAsnAsnSerGlyPheProSerAsnProLysPh 367
1092 TCGACAGAGAGGTGGCGGAACAACAGCGGCTTCCAGCAACCCCAATT 1141
367 eTrpLeuArgValSerGlnProSerGlnValTyrIleAlaValLeuGln 384
1142 CTGGCTGGCGGTCTCAAAACGAGTGAAGTGTACATTGCCCTCTGACA 1191
384 rSerArgLeuHisAlaAlaAspTrpAlaGlyArgAlaArgAlaLeuVal 400
1192 GATCCAGGCTGCACGCGGCGACTGGGACAGCGGGCGGCGGACAGTG 1241
401 GLyAspSerHisThrSerTrpSerProAlaSerIleProGlyLysHis 417
1242 GGTGACAGTCAATCTGTGTGAGCCAGGAGCATCCCGGCAAGCACTA 1291
417 rGlnAlaValGlyLeuHisLeuTrpLysValGlnLysArgArgValAsn 434
1292 CCGGGCTGTGGGTCTGCACCTCTGGAG..... 1319
434 euProArgValLeuSerMetProProValAlaGlyThrAlaCysHisAla 450
1319 ..... 1319
451 TyrAspArgGlnValHisLeuArgCysGlnLeuSerProGlyTyrTyrIle 467
1319 ..... 1319

```

```

467 uAlaValProSerThrPheLeuLysAspAlaProGlyGlnPheLeuA 484
1319 ..... 1319
484 rValPheSerThrGlyArgValSerLeuSerAlaIleArgAlaValAla 500
1319 ..... 1319
501 LysAsnThrThrProGlyAlaAlaLeuProAlaGlyGlnTrpGlyThr 517
1319 ..... 1319
517 lGlnLeuArgGlySerTrpArgValGlyGlnThrAlaGlyLysArg 534
1319 ..... 1319
534 snPheAlaSerTyrProThrAsnProCysPheProPheSerValPro 550
1319 ..... 1319
551 GlyProGlyProArgCysValArgIleThrLeuHisGlnHisCysArg 567
1319 ..... 1319
567 oSerAspThrGlnPheHisProIleGlyPheHisIlePheGlnValPro 584
1320 ..... GTCCAG 1326
584 lGlyGlyArgSerGlnAspAlaProProLeuLeuGlnGlnProLeu 600
1327 AGGTGGAAGAGCAGACGACGACCCCACTGCTGCTGAGAGCCGCTG 1376
601 LeuSerCysValProHisArgTyrAlaGlnGlnValSerArgLeuCys 617
1377 CTGAGCTGCTGCCACATCCCTACGCCAGAGGTGAGCGGCTGCTGCT 1426
617 uLeuProAlaGlyThrTyrLysValValProSerThrTyrLeuProAs 634
1427 CTTGCTGTGAGGACCTTACCAAGGTTGTGCTCCACTTCACTGCGGACA 1476
634 hGlnGlyAlaPheThrValThrIleAlaThrArgIleAspArgProSer 650
1477 CAGAGGGGCTTACAGATGATGATGATGATGATGATGATGATGATGAT 1526
651 lIleHisSerGlnGlnLeuLeuGlnGlnPheLeuGlnGlnValSerVal 667
1527 ATTCAACAGCAGAGAGATGCTGGGCAAGTTCCTCCAGAGAGTCCGTCAT 1576
667 tAlaValMetLysThr 672
1577 GGCAGTGATGAAAAAC 1592

```

seq_name: /SIDS1/gcgdata/geneseq/geneseqn_emb1/NA2001A.DAT:AAH17182

seq_documentation_block:

ID AAH17182 standard; cDNA; 1534 BP.

AAH17182;

26-JUN-2001 (first entry)

Human cDNA sequence SEQ ID NO:16542.

Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

Homo sapiens.

EP1074617-A2.

07-FEB-2001.

28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX (HELI-) HELIX RES INST.
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX WPT; 2001-318749/34.
 DR
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs
 XX
 PS Claim 8; SEQ ID 16542; 2537pp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 SQ Sequence 1534 BP; 286 A; 467 C; 507 G; 274 T; 0 other;

alignment_scores:
 Quality: 2487.50 Length: 511
 Ratio: 5.248 Gaps: 1
 Percent Similarity: 92.759 Percent Identity: 92.759
 alignment_block:
 US-09-768-877-2 x AAH17182
 Align seg 1/1 to: AAH17182 from: 1 to: 1534
 144 AspValPheTrpLeuProLeuLeuGluLysValTyrAlaLysValHisG1 160
 2 GATGGTTCCTGGCTCCCTTACTGAAAGAGCTTACGCCAAGGTCATGG 51
 160 ySerTyrGluHisLeuTrpAlaGlyGlnValAlaAspAlaLeuValAspL 177
 52 GTCTTACGAGCAGCTGTGGCGCGGAGGTGGCGGATGCCCTGTGGTGG 101
 177 euThrGlyLeuAlaGluArgTrpAsnLeuLysGlyValAlaGlySer 193
 102 TGACCGCGGCTGGCAGAAAGATGGAACCTGAAGGGCGTAGCAGGAAGC 151
 194 GlyGlyGlnGlnAspArgProGlyArgTrpGluHisArgThrCysArgG1 210
 152 GGAGCCAGCAGGACAGGACAGGCGCTGGGAGGACAGGACTTGTGGCA 201

210 nLeuLeuHisLeuLysAspGlnCysLeuIleSerCysCysValLeuSerP 227
 202 GCTGCTCCACCTGAGGACCACTGTGTATCAGCTGCTCGTGTCTAGCC 251
 227 roArgAlaGlyAlaArgGluLeuGlyGluPheHisAlaPheIleValSer 243
 252 CCAGAGCAGGTGCCCGGAGCTGGGGAGTTCATGCTTCATTTGCTCG 301
 244 AspLeuArgGluLeuGlnClyGlnAlaGlyGlnCysIleLeuLeuLeu 260
 302 GACCTGCGGGAGCTCAGGGTCAGGGCCAGTGCATCTCTGCTGTGG 351
 260 gileGlnAsnProTrpGlyArgCysTrpGlnGlyLeuTrpArgGlu 277
 352 GATCAGAACCTCGGGCGCGGTGTGGCAGGGCTCTGGAGAGAGG 401
 277 lyGlyGluGlyTrpSerGlnValAspAlaAlaValAlaSerGluLeu 293
 402 GGGGTGAAGGTGGAGCCAGGTAGATGCAGCGGTAGCATCTCAGCTCT 451
 294 SerGlnLeuGlnGlyGluPheTrpValGluGluGluPheLeu 310
 452 TCCAGCTCCAGGAAGGGAGTCTGGGTGGAGGAGGAGTCTCTCAG 501
 310 gGluPheAspGluLeuThrValGlyTyrProValThrGluAlaGlyHisL 327
 502 GGAGTTTACGAGCTCACCGTTGGCTACCGGTCTCAGGAGCGGCCACC 551
 327 euGlnSerLeuTyrThrGluArgLeuLeuCysHisThrArgAlaLeuPro 343
 552 TGCAGAGCTCTACACAGAGAGGCTGCTGCATACGCGGCGCTGCT 601
 344 GlyAlaTrpValLysGlyCysSerAlaGlyGlyCysArgAsnAsnSerG1 360
 602 GGGGCTGGTCAAGGCCAGTCAAGAGGCTGCCGGAACACACAGCGG 651
 360 yPheProSerAsnProLysPheTrpLeuArgValSerGluProSerGluV 377
 652 CTTTCCAGCAACCCAAATCTGCTGGGGTCTCAGACCGAGTGGAG 701
 377 alTyrIleAlaValLeuGlnArgSerArgLeuHisAlaAlaAspTrpAla 393
 702 TGTATATTGGCTCTCGAGAGATCCAGGCTGCAGCGCGGAGCTGGCA 751
 394 GlyArgAlaArgAlaLeuValGlyAspSerHisThrSerTrpSerProAl 410
 752 GCGCGGCGCGGCACTGGTGGTGACAGTCACTTCTGAGGAGCCAGC 801
 410 aserIleProGlyLysHisTyrGlnAlaValGlyLeuHisLeuTrpLysV 427
 802 GAGCATCCCGGCAAGCACTACCGAGTGTGGGTCTGCACCTCTGGAAGG 851
 427 alGluLysArgArgValAsnLeuProArgValLeuSerMetProProVal 443
 852 TAGAGAGCGGGGTCAATCTGCTAGGTCTCTGCTCCATGCCCGCCGTG 901
 444 AlaGlyThrAlaCysHisAlaTyrAspArgGluValHisLeuArgCysG1 460
 902 GCTGGCACCGGTGCCATGATACAGCGGAGGTCCACCTGCGCTTGTGA 951
 460 uLeuSerProGlyTyrTyrLeuAlaValProSerThrPheLeuLysAspA 477
 952 GCTCTACCGGGCTACTACCTGGCTGTCCCGAGCACCTTCTCTGAGGAGC 1001
 477 laProGlyGluPheLeuLeuArgValPheSerThrGlyArgValSerLeu 493
 1002 CGCCAGGGAGTCTCTGCTCCGAGTCTTCTACCGGGCGAGTCTCCTT 1051
 494 SerAlaIleArgAlaValAlaLysAsnThrThrProGlyAlaAlaLeuPr 510
 1052 AGCCGCCATCAGGGCAGTGGCCAGAACACACACCCCGGCGGAGCCCTGCC 1101
 510 oAlaGlyGluTrpGlyThrValGlnLeuArgGlySerTrpArgValGlyG 527

CC polypeptides and nucleic acids, and using these modulators to treat
CC diabetes, in particular through the regulation of an insulin
CC secretory response or insulin mediated glucose transport.
XX
SQ Sequence 49136 BP; 10300 A; 13118 C; 14165 G; 11553 T; 0 other;

alignment_scores:
Quality: 1989.00 Length: 3020
Ratio: 3.224 Gaps: 11
Percent Similarity: 20.430 Percent Identity: 20.298

alignment_block:
US-09-768-877-2 x AAA27475

Align seg 1/1 to: AAA27475 from: 1 to: 49136

47 GlnGluileCysAlaThrProArgLeuPheProAspProArgGluG1 63
|||||
3810 CAGGAGATTGTGCCACACCCCGCTGTTCCAGATGACCCAGGGAAGG 3859

63 yGlnVallyGlnGlyLeuGlyAspCysTrpPheLeuGlyAlaCysA 80
|||||
3860 GCAGGTTAAAGCAGGGGCTGCTGGGGATTGCTGCTCTCTGCTGCG 3909

80 laAlaLeuGlnLysSerArgHisLeuLeuAsp..... 90
|||||
3910 CCGCGCTCAGAGAGCAGGACCTCTCGACAGAGTGGGGGCCCTTTC 3959
90 90

3960 CCTGTGTTGTCCTCGAGCGGGTTCCTTTTTCCTCCAGCCCTGCTG 4009
90 90

4010 AGTACCAGGAGGCTTTCGAAAGCAGAGCTGTGCCGACGCCGATCTCT 4059
90 90

4060 GCTGTGTTGGGGAAGGAGGAGAGAGTTCACAGGACAGAGCTGAGGACTGC 4109
90 90

4110 ACTGTCTCCTCTGCTGACGGGGGGTGGCTTGGCTGCCAGAGAGGCTC 4159
90 90

4160 CATCAGGAGGTTCCGCTCTCTGTGCTCTCTGACCCCGGACTCCAT 4209
90 90

4210 GGAGTCAGATCACCAGGTTTGAATAAAGAGACAAATGTGCCAGCTACA 4259
90 90

4260 GGAGGAGGGGCTGCTGCGACGCTCTGCTCAGATCTCTCTCAGCTAG 4309
90 90

4310 CTCGCTGCTTTTCTTACAGGTTTGAATAAAGTTCGCAAAAGTTATTA 4359
90 90

4360 AACCTGTTTCTGTTGGGTAGACAGATATCTTGGGAGGAGAGGCTTCTCA 4409
90 90

4410 GGTTCCTTACCTGGGAGTGTTCACCGTTTATGCTTGGCTTGTGCTA 4459
90 90

4460 AGTGTGCTGATTAAATGACAGCGGCTCAACAGTGTGACCTCATTTCAGAGT 4509

FT exon 30857..30980
/*tag= ad
/number= 14
FT intron 30981..31445
/*tag= ae
/number= 14
FT exon 31446..32175
/*tag= af
/number= 15
FT variation replace(6225,a)
/*tag= ag
/frequency= 0.75
/note= "G-allele has a frequency of 0.75 in Mexican Americans"
FT /frequency= 0.71
/note= "G-allele has a frequency of 0.71 in non-Hispanic whites of German ancestry"
FT /frequency= 0.90
/note= "G-allele has a frequency of 0.90 in African Americans"
FT /frequency= 0.94
/note= "G-allele has a frequency of 0.94 in Asians (Japanese)"
FT CDS 43070..44944
/*tag= ah
/note= "encodes GPR35 (AA79576)"
FT polyA_signal 44922..44927
/*tag= ai
FT PN W0200023603-A2.
XX 27-APR-2000.
PD
XX 21-OCT-1999; 99WO-US24890.
XX
XX 21-OCT-1998; 98US-0105052.
PR 13-MAY-1999; 99US-0134175.
PR
XX (ARCH-) ARCH DEV CORP.
XX
XX Polonsky KS, Horikawa Y, Oda N, Sreenan S, Zhou Y, Otani K;
PI Hanis CL, Bell GI;
XX
XX WPI; 2000-339702/29.
DR P-ESDB; AAY79567, AAY79568, AAY79569, AAY79570, AAY79571, AAY79572,
DR AAY79573, AAY79574, AAY79576.
XX
XX Method for screening for type 2 diabetes mellitus comprises detecting a
PT polymorphism in a calpain encoding nucleic acid segment or a
PT protease-encoding nucleic acid segment -
XX
XX Claim 65; Page 203-217; 257pp; English.
XX
XX The present sequence is that of a 49,136 bp region located within
CC the NIDDM1 region of human chromosome 2. It includes the CAPN10
CC gene that encodes a novel calpain-like cysteine protease, designated
CC calpain 10, and a gene encoding a G protein coupled receptor,
CC GPR35. Alternative splicing of calpain 10 mRNA generates a family
CC of proteins. Isoforms 10a-h (see AAY79567-74) are respectively
CC encoded by exons 1-7,9-13, 1-7,9,10*,11-13, 1-7,11-13, 1-7,9,11-13,
CC 1-10*,11-13, 1-3*,4-7,9-13, 1,2,14,15 and 1,11-13. Calpain 10 mRNA
CC is ubiquitously expressed; the major 2.7 kb transcript was detected
CC in every human adult and foetal tissue examined. Mutations in the
CC CAPN10 gene are responsible for susceptibility to type 2 diabetes.
CC The nucleotide variant showing all the evidence for linkage to
CC type 2 diabetes, UCSNP-43, is located at nucleotide 6225 in intron
CC 3 of the gene. There is alternative splicing of intron 3, but the
CC molecular mechanism by which the polymorphism at UCSNP-43 affects
CC susceptibility to type 2 diabetes is unclear. Claimed methods for
CC screening for a propensity for type 2 diabetes mellitus are based
CC on detection of a polymorphism in a calpain encoding nucleic acid,
CC especially UCSNP-10 of the CAPN10 gene. Methods are also claimed
CC for identifying modulators of calpain activity using calpain 10

90 90
4510 TTCACATATGCCAGGCCCATGTAAAGCTGTCACAGTCACTGGCTTT 4559
90 90
4560 CAGACATATGCTTACCACTTGTACTTTTAAAAAGAGACTGCTA 4609
90 90
4610 AAATCACTGCCATTGTGTTCTGAGCCGTAAGTGCACAGACGACGACCGC 4659
90 90
4660 ACCAGTCCCGTGCCAGACCTGCACCTCCCGATGCTGGTCAGAATGC 4709
90 90
4710 TTACCCCTGAAGAGCCCTCGGCTGACGCTGTGGGTCAAGACAGTGGC 4759
90 90
4760 CCAGAGTCGGGCGCCAGAGCTCCACAGCAGAGAGGGGCTGCTGCTCTG 4809
90 90
4810 TGTGAGCTGTGCTTGACGCCAGCTCGTCAAGAACTGGGTCACTCATG 4859
90 90
4860 CCCTGATATGCATTTGTTGGCTTCAGGTCAGATGCTTTAGTGAAG 4909
90 90
4910 GGCACAGAGTGTGCCCGGATATGTGGCTCCCTCGGTGTGCTCTCA 4959
90 90
4960 GTTTGCAATGAGAGGTCTGTTAATTCATGTGGTATGACGCCCTGTG 5009
90 90
5010 CAGCGCCGACATCCAGSTGTGCGTAGAGTCTCTCGACATCCAGGTG 5059
90 90
5060 TGCCGTAGAGTCTCTGCGACATCCAGGTGTGCCGTAGAGTCTCTGCGA 5109
90 90
5110 CATCCAGGTGTGCCGTAGAGTCTCTGCGACATCCAGGTGTGCGTAGAG 5159
90 90
5160 TTCTCTCAGACCGCGGTGCTGTGAGCACTAGCTGTGCGCACACCGC 5209
90 90
5210 GGCCGGACACATGAGTAGCGCGGTGTATATATCAAGGCTCGGCTT 5259
91 91
5260 TTGCTTCTCCCTGTGATGGCAGGTCAATCTCCGCGACAGCCGAGCTGG 5309
101 AlasplnglnlurYrarglySerPheThrCysArgIleTrrpGlnPheG 117
|||||
5310 GCCGACCAAGAGTACCGGGGCTCTTCACTGTGCGCATTTGGCATTTGG 5359
117 YArGTPvAlgluValThrThraspAspArgLeuProCysLeuAlaGlyA 134
|||||
5360 ACGCTGGGTGGAGTGAACACAGATGACCGCTGCGCTTGCACAGGA 5409
134 rglaucyspheserArgCysGlnArgGlnAspValPheTrrPleuProLeu 150

|||||
5410 GACTCTGTTTCCCGCTGCCAGAGGAGAGATGCTTCTGCTCCCTTA 5459
151 LeuGlnLysValTyrAlaLys..... 157
5460 CTGMAAAGGTCTACGCCCAAGTGCCTGTGTGGGGCTGAAGGGCTGGC 5509
157 157
5510 CTGGGGCAAGTGGAGCTGCCATACATAGGAGTGCCTCCAGAGAGTCTC 5559
157 157
5560 TGTCACTCTGGGCTGCAGAGCCCTTCACTTCTGAGAGTCTGGCACT 5609
157 157
5610 CATTCGTGACTCAGGCTGACAGGCCAGTGCAGAGATTCYTTTGGG 5659
157 157
5660 CTTGTGATTGCCACATCCCTGCTTCCCTTCCCTTCCAAAGCCAG 5709
157 157
5710 CGTGAGTCTGTTCTCAAGAGAAATGTGTGCCGTCTCTTATTTAT 5759
157 157
5760 CGGCCCAAGAAAGATGCTCTTATATTGTGTGAAATGTGGG 5809
157 157
5810 ACAGGCACTCATTTGTGTACTGCTGGGAGAGTAGGCTACCCAGC 5859
157 157
5860 ATACCAACTTGTGTATCAAGGTGCTTGTGGCTCAGGGGACCAAGACC 5909
157 157
5910 CTCACATGATCATTAATTGAATAGCCTTCCCTTAGAATGATTTGTC 5959
157 157
5960 TTCTTCCAAAGCACTGACTGACTACAGGACGAGGAAAGCTGTGAAC 6009
157 157
6010 ATGGAGAGCTGGCTGTGACATCATGCCCAGTAGACCTTCCATCCA 6059
157 157
6060 AGGCTGTTTTAGAAAAGCAGGTTGAGCTTGAGACCAAGGATGTG 6109
157 157
6110 GGCATCATATAGCTTCCAGGCTCTCTGCTGCTCTGTGCCACACCGGA 6159
157 157
6160 TGCCAGAGACTTCTGTGTGGCAGAGACTGACAGGGGCTCACGCTT 6209
157 157
6210 GCTGCAAGTAAGCGTTTGAAGTGTAGGCTAAGCCTTGACTTGGTAGG 6259
157 157
6260 ATAGAGAAAGCAGAGGGGAGTAAAGAGTGGATTTAGGACAGCGTT 6309
157 157

6310 GGAGGATTGGGGTGCTACAGACCATTGGGAATCAGAGAGGGGGCCATGCT 6359
158ValHisgl 160
6360 CAATGCCAGAGGCTCACTCCCATGGTGATTGTGTCCCTAGGGTCCATGG 6409
160 ySerTyrGluHisLeuTrpAlaGlyGlnValalaAspalaLeuValAspL 177
6410 GTCCCTACGAGACACTGTGGCGGGCGAGGTGGCGGATGCCCTGGTGGACC 6459
177 euThrGlyGlyLeuAlaGluArgTrpAsnLeuLysGlyValalaGlySer 193
6460 TGACCGCGCGCTGGCAGAAAGATGGAACCTGAAGGGCGTAGCAGGAAGC 6509
194 GlyGlyGlnGlnAspArgProGlyArgTrpGluHisArgThrCysArgG 210
6510 GGAGGCCACGAGACAGCCAGCGCGCTGGAGCACAGACTGTGCGCA 6559
210 nLeuLeuHisLeuLysAspGlnCysLeuLysSerCysValLeuSerP 227
6560 GCTGTCCACCTGAAGGACCAGTGTCTGATCAGCTGCTGCTGCTCAGCC 6609
227 roArgAlaGly 230
6610 CCAGAGCAGGTGAGGCACGTGGCCCAACATGGGAGGGCTGCAGCCAGCGTG 6659
230 230
6660 CCCCCCACTGCCAGGCTCAGGCACACTGTAGCTTTTATGTACTGGCT 6709
230 230
6710 ACACAGCCCTGTAGGACTTAAGTGGGAAGAGTAAGCTTGTCTCAAGGG 6759
230 230
6760 TGGTGTCTCACTTTGTGACCTTCCCTGCTGCTCTCTTCCAGAGGGAGC 6809
230 230
6810 TGGCCCTTCTCTCCCTGACCACTCTTCCACTAGTGCAGGCGAGGAAG 6859
230 230
6860 AGGTGGCACCGAGTCAAGGCCCACTGTGTGCCATCCCTGGCCCACTG 6909
230 230
6910 GCAACTGGCAAAATCAAAACCTGTTTTTTATTTAGTGATAGATAACAT 6959
230 230
6960 TCGTTAAAAACAGTTTGTCTCCAAAAAANTGAAAGGGCGCAGGTGGTGG 7009
230 230
7010 CTCACACCTGTAATCCAGAACTTTGGGAGGCTGAGTGGGAGTATCGCT 7059
230 230
7060 GGAGCCAGGAGTTCAAGAGACCCCGCTGGGCAACATGGCAAGATCTCAT 7109
230 230
7110 CTCACAAAAAATGAAGAAAAAANTCACTTAGATGGAACACATGTGAC 7159
230 230
7160 TTTTGTAGTGCCTCTCAGTTTTTCCATGAGCAGCAGCGTTTACGTGTCC 7209
230 230
7210 CTTCCGACCGCTTCTCACACTGCCACACACGCGCTGTCAAAATGTCCGC 7259

230 230
7260 CCATGAGCGGGTTTCCACAGTCTCTTAATCATTCCTCGATGTTGGATA 7309
230 230
7310 TTAAGATCTCTCTGTTTTATATATATTATAAGCAGTATCAATGAACATCT 7359
230 230
7360 TTATCATTTTTTTCATACTTAGGATTATTTAAAAATATGGTTAAAGAAT 7409
230 230
7410 ATGAATATCACAGTAAGTAAGTGAACAGTTGTCAATAGGCCCTTGGTCTGGGT 7459
230 230
7460 CCCCCAATAAGCAGACCCCTGAGATGAGTTTCAGGACACGTTTCAGAGGTTTC 7509
230 230
7510 AGAGAGCCTGGAGGGCGTGTACTCCCCCGCAGCCCCGCTGGGGTGCAGG 7559
230 230
7560 AAGGACGCTGGCCTCAGAGCCTCCACCTGGGAGTGAAGGGCGCTGCTGTG 7609
230 230
7610 GTGGSCATTAGCTCGGGGAGCTGTTGTTTTGGGTGCTCCAGGGTGGGGT 7659
230 230
7660 AGCGCTAAGTCTTAGCACTTCAGGCTTTAGAGAAAGCCCCCAGGCAGAGA 7709
230 230
7710 GAGATGGCAGCTGGCAGTGAAGTGGAGTGCATTTGGAGCCTGCTGAGGTGG 7759
230 230
7760 CAAGGGCGCGGCTGGCGGAGACTGAAGTCCCTTCCAGAGAGACCTCTCT 7809
230 230
7810 CTCAGACCGTTGCCCGCCACAGATGTGGCTTCTCTCTGTTTGGGCTC 7859
230 230
7860 TTTCTTTTCTATGCCCCATCTGTCTCAAGGTCTCTGGCCAGGTGGAGCTGT 7909
230 230
7910 TGCCCCCTGGGCTCCATGCTCCAGTGGGCTTCTAGCCAGGCACATAT 7959
230 230
7960 CTGAATGTCTTACCTCTAGCAGGCTTTGCCAAGAGAAACACTCTCTAGAT 8009
230 230
8010 TTCTGGCAGTTGCAGATGCATCTGTGCAGCATGTTTGAGAACTGGTGTGCTG 8059
230 230
8060 TGCCAGGCATCTGCACAGATGGGTACGGAGATGACTGAGGCCCTAGAGA 8109
230 230
8110 AGGGGATGACTTACCAGCACAGACCCGGATGGCAGCCGAGTCAAGGCC 8159


```
230 ..... 230
8160 CGTGCTGCTGCTCGAGATGCTCTCTAGCTGATGAGGTGACAGCTGCTA 8209
230 ..... 230
8210 AGAAGAGACTCTCGTGTGATGGGGAATCTGGGGTCTCCCTGTAGCCA 8259
230 ..... 230
8260 TAGTGGGTGGGCTGGAGCATCTCCAGAGAGAGCAGAGGCTGTGTG 8309
230 ..... 230
8310 TCCTTGTAGTTGGGACTCCATGTGCTCCCTGCTGCTGCTGCTGCTG 8359
231 .....Ala..ArgGluLeuGlyGluPheHisAlaPheIleVal 242
8360 ATTCTCATGAGAGTCCCGGAGAGCTGGGGAGTTCATGCTTATTTGT 8409
242 1SerAspLeuArgGluLeuGlnGlyGlnAlaGlyGlnCysIleLeuLeu 259
8410 CTCGGACCTGGGGAGCTCCAGAGGTTCAGCGGGCCAGTGCATCTGTGTC 8459
259 euArgIleGlnAsnProTrrpGlyArgArgCysTrrpGlnGlyLeuTrrpArg 275
8460 TGGGATCCAGAACCCCTGGGGCCGGGGTGTGGCAGGGGCTGTGGAGA 8509
276 GluGly..... 277
8510 GAGGGGTAGTGTGGGCTGGGACCATGTGCTGGGAGGGGGGCC 8559
277 ..... 277
8560 AGTCCAGCTGGGCTGTGCTCTGTGTCATCTACGTCTGACGACTGTAC 8609
277 ..... 277
8610 TTGGCTGTCTCCAGAGAGCCCTGATGCTCTGCTGTGACACCATGC 8659
277 ..... 277
8660 TTGTCTGGCTCCAGAGCAATCTGTGTGAGGCTGGGACCAAGTGGCCAT 8709
277 ..... 277
8710 TGGGCTGGGGTTTCATATGAGGAGACATCATCAGGGCTGGGACAGAG 8759
277 ..... 277
8760 TCTCGGAGAGAGCATCCAGAGAGGCTGAATTCCTCGGAAAGAGAGG 8809
277 ..... 277
8810 GCTCAGAGGTGGCCGTGCCAGAGAGGCTGTGCTGAGAGCTGTCTGG 8859
277 ..... 277
8860 GGTGTGGAGGGGCTGCAGAGCTGGGGCACGGGTGCTGTGACAGAACTC 8909
277 ..... 277
8910 AGGGCTCTTGGGTCCCTCCAGGCTTCCCCAGCCTGCCGGCAAGTTG 8959
277 ..... 277
8960 ACACTACAGTTCTCGGGAGGGGCTTCTGCTGAGATGAGGTTTCTTCAG 9009
278 .GlyGluGlyTrrpSerGlnValAspAlaAlaValAlaSerGluLeuLeu 294
9010 GGGTGAAGGGGTGGAGCCAGTAGATGACGGGTAGCATCTGAGCTCTGT 9059
294 erGlnLeuGlnGlyGlyGluPheTrrpValGluGlnGluGluPheLeuArg 310
|||||
9060 CCCAGCTCCAGAAAGGGAGTTCTGGGTGGAGAGAGAGAGTCTCTCAGG 9109
311 GluPheAspGluLeuThrValGlyTrrpValThrGluAlaGlyHisLe 327
9110 GAGTTTGACAGAGCTCACCGTTGGCTACCCGGTACAGGAGGCCGC.CACTT 9158
327 uGlnSerLeuTrrpThr..... 332
9159 GCAGAGCCTTACACAGTAGTCCCGCAGAGGGGCTGTGCTGGGACAGTGC 9208
332 ..... 332
9209 TCTGCTCCGAGTAGAGAGAGGCTGGCAGGCTGCTGGTTCCCTGTC 9258
332 ..... 332
9259 CCAGGCCAGTTGGTTGTTCTCTACAGCTGAGAGATGATCTGTCCAGG 9308
332 ..... 332
9309 AGCCGGAGAGAGGATGATGATTTCTGCCAGAGACTGGAGAGAGGTGG 9358
332 ..... 332
9359 CTTGTGGAGGGGCTGCTGTGCTGTGCGGTAGCTGCTGCTTAGACCC 9408
332 ..... 332
9409 TGCAGAGTTATGAGGCCACCTGGGGAGGCCAGGAGAGCCGTGT 9458
333 .....GluArgLeuLeuCysHisT 339
9459 CCCAGAGCTGATGCTGTGTTTCTCTACATAGAGAGGCTGTGCTCCATA 9508
339 hrArgAlaLeuProGlyAlaTrrpValLysGlyGlnSerAlaGlyLys 355
9509 CGCGGGCTCTCCCTGGGGCTGGGGTCAAGGGCCAGTGCATGAGAGCTGC 9558
356 ArgAsnAsnSerGlyPheProSerAsnProLysPheTrrpLeuArgValSe 372
9559 CGGAACACAGCCGCTTTCAGCAACCCCAAAATTCGCTGCGGGCTC 9608
372 rGluProSerGluValTrrpIleAlaValLeuGlnArgSerArgLeuHis 389
9609 AGAACCGAGTAGGTATGATTCGCTGCTCCAGAGATCCAGGCTGCACG 9658
389 1aAlaAspTrrpAlaGlyArgAlaArgAlaLeuValGlyAspSerHisThr 405
9659 CGCGGAGCTGGGACAGGCCGGGCCGACACTGTGGGTGACATGACTACT 9708
406 SerTrrpSerProAlaSerIleProGlyLysHisTrrpGlnAlaValGlyLe 422
9709 TGTGGAGCCCGAGCGAGCATGCCGGCAAGCATACAGGCTGTGGGTCT 9758
422 uHisLeuTrrpLysVal..Glu..... 428
9759 GCACCTCTGAGAGTAACTCAAGCCCGTGTGCTCAGCTGTGGTTCAGCA 9808
428 ..... 428
9809 GGTGTGTGAGGCCCATGAGAGTCTGGGTCTAGAGCTGGCTGTGCCGG 9858
428 ..... 428
9859 GACACATGACTCTGTCCACGGGGCCACAGTCTCCCTCCTTGGGC 9908
428 ..... 428
9909 TGTTCACAGGGGTGACGTGTGCTGTGCTGCCAGACCCGGCTGTGACT 9958
428 ..... 428
```

9959 GAGACTGCAGTCTTTCTGCTTCCCGTGTGCTCATTTGGCCAAAGGAAA 10008
428 428
10009 GCACAGAGTCTGCAGCCAGGACCGCAGGAGGGCCTGGACCCG 10058
428 428
10059 GGGGCTCTTGGCAGCGCGTCTTCTGAGCAAGAGGTAGAGCCAG 10108
428 428
10109 CGGCTGAGGACCTGTACAGGCCAGTCCAGCTGTGCAGTGTGTGTGA 10158
428 428
10159 CCTGGCACACATCTCTCCCTGCTCCTCAGTCTTTCCTCCCTGCAAGAC 10208
428 428
10209 GGGTCTGTACAGGATCTCATGGATTGCTGTGAGGCCAGGAGTCCC 10258
428 428
10259 AGGCTAACCACTGGTTCAAAAGTGTGTTTCCAGGAAGAACAGATG 10308
428 428
10309 GGGGCGCTCAGGGCAAGGCGCTGAGTGTGTCGAGGATATCCCGCTG 10358
428 428
10359 CTCGCTCAGGGGCTGGTTTTCATCTTGTGTGTGACAGGTTGTACA 10408
428 428
10409 CTTGGACCACACTGTTCCCTGTCCCTTCATGATGTGGCCACATGATG 10458
428 428
10459 TTCCTTTCTCTTGAAGAAGTGTGTGAAGGCCCACTGTCCAGCAGC 10508
428 428
10509 CCCCAGTTGCTTGGCCACAGTGTGCTTGTGGGCCAGCTACAGGAGG 10558
428 428
10559 ACTTGCAAGCTGTGCTGGGACAGATGCTGGGCCAGGGCCAAAGTGAAG 10608
428 428
10609 CCGGGATTGTTGGGCATCTCTAGTGTGCTCCTGAGAGAGGTGGAGGT 10658
428 428
10659 GTTGACAGGCTTGGCGCTTTTCATGTCAACTCCAGAGGCCCTGTGTCT 10708
428 428
10709 TGCAGCAGGAGGTCAAGGCCAGGGCTGTGACCCCGCGCTCTCCAC 10758
429
10759 ACTGAGCTCTGCAGTGTCTACAGGTAGAGAAGCGCGGTCAATCTG 10808
435 ProArgValLeuSerMetProProValAlaGlyThrAlaCysHisAlaTy 451
10809 CTTAGGGTCTCTTCCATGCCCGCTGGTGGTGCACCGTGTGCATGCATA 10858
451 rAspArgGluValHisLeuArgCysGlnLeuSerProGlyTyrTyrLeuA 468
10859 CGACCGGAGGTCCACCTCGTGTGTGAGCTCTCACCGGCTACTACCTGG 10908

468 laValProSerThrPheLeuLysAspAlaProGlyGluPheLeuLeuArg 484
10909 CTCTCCCAAGCAGCTTCTGAAGCAGCCGCGAGGGAGTCTCTGCTCCGA 10958
485 ValPheSerThrGlyArgValSerLeu..... 493
10959 GTCTTCTTACCGGGCAGTCTCCTTAGGTAGAGGAACCGCCAGTGC 11008
493 493
11009 TGCTGGCTCTCCGAGGCCACAGGCCCTTCCAAGGCAGGATTGGGCACCT 11058
493 493
11059 TCCTCTGTGTTGGCAGGTGTCCATGTGGAACTGAGGCCACTTGGGAAC 11108
494 SerAlaIleArgAl 498
11109 CTGCTGCCAGGCCCTCCCATGTTTGTCTTCTTGGCAGCCCATCAGGGC 11158
498 aValAlaLysAsnThrThrProGlyAlaAlaLeuProAlaGlyGluTrpG 515
11159 AGTGGCAAGACACACCCCGGGCAGCCTGCTCGGGGAGTGGG 11208
515 lyThrValGlnLeuArgGlySerTrpArgValGlyGlnThrAlaGlyGly 531
11209 GGACCTGTGAGTACGGGGTCTTGGAGAGTGGCCAGAGCGGGGGGC 11258
532 SerArgAsnPheAlaSerTyrProThrAsnProCysPheProPheSerVa 548
11259 AGCAGGAACCTTGGCTCATACCCACCAACCCCTGCTTCCCTCTCGGT 11308
548 lProGluGlyProGlyProArgCysValArgIleThrLeuHisGlnHisC 565
11309 CCGCAGGCGCTTGGCCCGCTGCTGCTGCCATCACTCTGCATCAGCACT 11358
565 ysArgProSerAspThrGluPheHisProIleGlyPheHisIlePhe... 580
11359 GCGGCCAGTGCACCCAGTTCACCCCATCGCTTCCATATCTTCCAG 11408
580 580
11409 GCAAGCTCTTCCCGCAGGAGGGGGGAGGAGGAGGAGGAGGAGGAGG 11458
580 580
11459 AATTGTCATTTGGCTTCCATTTGCCAACAGAGGCTCTGGGCTCAGTC 11508
580 580
11509 ACTTGGGCTCCCTGCTTCCCTTCCAGGCGCTGCTAGAACCCGACAGGGC 11558
580 580
11559 CCTCTCCATCTCCAACCTCTCAGAGGCAAGCCGAAGATGGCTCTGGA 11608
580 580
11609 AGGGCCGGGGCTGGAGGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAG 11658
580 580
11659 TCCAGAGGAGGTGGTGTGGGGAGGAGGAGGAGGAGGAGGAGGAGGAG 11708
580 580
11709 GAGGCCAGGTTGAGACAGCGGGGTGGTGGGAGGAGGAGGAGGAGGAG 11758
580 580
11759 GCAGGGGCTGTGACTGAGGAGTGCAGAGGAGGAGGAGGAGGAGGAGGAG 11808

```

580 ..... 580
11809 TGGCCACAGACCCCTGGGTGTGAGAGAGTCCACAGTCACTGAGTCA 11858
580 ..... 580
11859 GCCACAGACCCCTGTCCACTTACCTGACTGAGAAATGACTGTCCAG 11908
580 ..... 580
11909 GTTATTCTCTGACAGATGTGCCCTGGAATGAGGGGCTGACGAG 11958
580 ..... 580
11959 GAAGGACATGACACCTCGGTTACAGTGGGCTGGGGAGCCCTTGA 12008
580 ..... 580
12009 CCTCGCTGTTTGGCTGGGCGGCGTCAAGTCCCTAGAGCTTGAG 12058
580 ..... 580
12059 GAAACACATGCCAGGCGCAGTGGAGCCCTTGGGGGGGCTGGGCAATC 12108
580 ..... 580
12109 ACAAGTGTAAAGCCCTGATGATGAGAGCCCTCAGGCGGGGCCCC 12158
580 ..... 580
12159 ACTGCCGACACTTCTGGCAGAGGCTGGCCAGGCTTGGTGAAGAGCGAG 12208
580 ..... 580
12209 TCCAGTGTCCAGGCTGGAGCCCTCTCAGAGAGGGGCTGATGTA 12258
580 ..... 580
12259 CTCAGAGGGCCAAAGGCAATCCAGAGATGGCTGGGCTCCCT 12308
581 ..... GlnValPr 583
12309 ACCCCAGGCTGGCCCCCTCAGTCTGAGCCCTTCTCAGGTC 12358
583 .....
583 ogtlglgIyArgSerGlnAspAlaProLeuLeuGlnIuprol 600
12359 AGAGGGTGAAGAGCAGAGACGCCACCTGCTCTCAGAGAGCCG 12408
600 euleuSerCysValProHisArgTyrAlaGlnIuValSerArgLeuCys 616
12409 TGTGTAGCTGTGTCACATGCTAGCCGAGAGGTGAGCCGCTGTC 12458
617 LeuLeuProAlaGlyThrTyrLysValIalProSerThrTyrLeuProAs 633
12459 CTCTCTCTGCGGGCAGCTTACAAAGTTGTCCCTCAGCTACCTGCCGA 12508
633 pThrlGlnIyAlaPheThrValThrIleAlaThrArgIleAsp..... 647
12509 CACAGAGGGGCTTTCACAGTGCATTCGACACAGAGATTGACAGGGG 12558
647 ..... 647
12559 GCTGTGGAGCTTGGGGGCGGCGCAGTGAAGGTGGGTCTGAGTCTTA 12608
647 ..... 647
12609 GTGCTGCTGTGCCCCCAGCTCTCTGCTGCTCCCTCAGCCCTCAAGCC 12658
647 ..... 647
12659 CTATCTGCTCTGGCAGACAGAGGCTCTCTGCTTACCTGAGGAGCCCTTC 12708
647 ..... 647

```

```

12709 TTGCTGTCTGAGCCTGAGAGAGAGTCTAGTGGAGGTGGGCCAGGAGC 12758
647 ..... 647
12759 ACACAGCCACTTGTGTGACAAAGTCACTGGAGCGCTGATGTGTCTC 12808
648 ..... ArpProSerIleHisSerGlnIuMelLeuGlnIuPheLeuG 662
12809 TCTCCACAGGCGCATTCATTCACAGCAGAGATGCTGGCCAGTCCCTCC 12858
662 InGluVal 664
12859 AAGAGGTG 12866

seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH04003

seq_documentation_block:
ID   AAH04003 standard; cDNA; 874 BP.
XX
AC   AAH04003;
XX
DT   26-JUN-2001 (first entry)
XX
DE   Human cDNA clone (5'-primer) SEQ ID NO:838.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999; 99JP-0248036.
XX
PR   27-AUG-1999; 99JP-0300253.
XX
PR   11-JAN-2000; 2000JP-0118776.
XX
PR   02-MAY-2000; 2000JP-0183767.
XX
PR   09-JUN-2000; 2000JP-0241899.
XX
XX
XX   (HELI-) HELIX RES INST.
XX
PI   Ota T, Isogai T, Nishikawa T, Hayashi K, Salto K, Yamamoto J;
PI   Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX   WPI; 2001-318749/34.
XX
PT   primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs.
XX
PS   Claim 1; SEQ ID 838; 2537pp + CD ROM; English.
XX
XX
XX   The present invention describes primer sets for synthesizing 5602
XX   full-length cDNAs defined in the specification. Where a primer set
XX   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX   to the complementary strand of a polynucleotide which comprises one of
XX   the 5602 nucleotide sequences defined in the specification, where the
XX   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX   of an oligonucleotide comprising a sequence complementary to the
XX   complementary strand of a polynucleotide which comprises a 5'-end
XX   sequence and an oligonucleotide comprising a sequence complementary to a
XX   polynucleotide which comprises a 3'-end sequence, where the
XX   oligonucleotide comprises at least 15 nucleotides and the combination of
XX   the 5'-end sequence/3'-end sequence is selected from those defined in
XX   the specification. The primer sets can be used in antisense therapy and
XX   in gene therapy. The primers are useful for synthesizing polynucleotides,
XX   particularly full-length cDNAs. The primers are also useful for the
XX   detection and/or diagnosis of the abnormality of the proteins encoded by
XX   the full-length cDNAs. The primers allow obtaining of the full-length

```

CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 874 BP; 161 A; 238 C; 315 G; 155 T; 5 other;

alignment_scores:
Quality: 1251.00 Length: 286
Ratio: 4.849 Gaps: 7
Percent Similarity: 90.210 Percent Identity: 86.364
alignment_block:
US-09-768-877-2 x AAH04003
Align seg 1/1 to: AAH04003 from: 1 to: 874

145 valPheTrpLeuProLeuLeuGluLysValTyrAlaLysValHisGlySe 161
|||||
5 GTCTTCTGGCTCCCTTACTGGAAGAGTCTACGCCAAGTCCATGGTGC 54
|||||
161 rTyrGluHisLeuTrpAlaGlyGlnValAla.AspAlaLeuValAspLeu 177
|||||
55 CTACGAGCCTCTGTGGCCGGCGAGGTGGTGGATGCTGTGGGACCTG 104
|||||
178 ThrGlyGlyLeuAlaGluArgTrpAsnLeuLysGlyValAlaGlySerGI 194
|||||
105 ACCGGCGGCTGSCAAGAGTGAACCTGAAGGGCTAGCAGGAAGCGG 154
|||||
194 yGluGlnGlnAspArgProGlyArgTrpGluHisArgThrCysArgGlnL 211
|||||
155 AGCCACGACAGGACAGACAGCGCGTGGGAGCACAGGACTTGTGGCAGC 204
|||||
211 euLeuHisLeuLysAspGlnCysLeuLysLeuSerCysCysValLeuSerPro 227
|||||
205 TGTCTCACTGAGGACAGTCTGTATCAGCTGCTGCGTGTACGCCCC 254
|||||
228 ArgAlaGlyAlaArgGluLeuGlyGluPheHisAlaPheIleValSerAs 244
|||||
255 AGAGCAGGTGCCCGGAGCTGGGGAGTTCATGCTTCCATGCTCTCGGA 304
|||||
244 pLeuArgGluLeuGlnGlyGlnAlaGlyGlnCysIleLeuLeuLeuArgI 261
|||||
305 CTGCGGGAGCTCCAGGGTCCAGCGGGCCAGTGCATCTCTGCTGCGGA 354
|||||
261 leGlnAsnProTrpGlyArgCysTrpGlnGlyLeuTrpArgGluGly 277
|||||
355 TCCAGAACCCCTGGGGCCCGCGGTGCTGGCAGGGCTCTGGAGAGAGGG 404
|||||
278 GlyGluGlyTrpSerGlnValAspAlaValAlaSerGluLeuLeuSe 294
|||||
405 GGTGAAGGCTGGAGCCAGGTAGATCAGCGGTAGCATCTGAGCTCTGTC 454
|||||
294 rGlnLeuGlnGluGlyGluPheTrpValGluGluGluPheLeuArgG 311
|||||
455 CCAGCTCCAGGAAGGGAGTCTTGGTGGAGGAGGAGTCTCTCAGGG 504
|||||
311 luPheAspGluLeuThrValGlyTyrProValThrGluAlaGlyHisLeu 327
|||||
505 AGTTTGACGAGCTACCGTGGCTACCGGTACCGGTACCGAGCGCCACCTG 554
|||||
328 GlnSerLeuTyrThrGluArgLeuLeuCysHisThrArgAlaLeuProGI 344
|||||
555 CAGAGCCTACACAGAGAGGTGCTCTGCCATACCGGGCGCTCGCTGG 604
|||||
344 yAlaTrpValLysGlyGlnSerAlaGlyCysArgAsnAsnSerGlyP 361
|||||
605 GGCCTGGGTCAAGGGCCAGTACGAGGAGTGTGCGGGAACAACAGCGCT 654
|||||
361 heProSerAsnProLysPhe.TrpLeuArgValSerGluProSerGluVa 377
|| |||||

655 TTNCAGCAACCCAAATCTTCTTGGCGGTCTCAGAACCCGAATGAGGT 704
377 lTyrIleAla...ValLeuGlnArgSerArgLeu...HisAlaAlaAspT 392
|||||
705 GTATCTGGCGGCTCTGTCANNAAGTCCAGGCTTGAACGCCGCGGACT 754
392 rp.....AlaGlyArgAlaArgAlaLeuValGly...AspSerHisThr 405
|| |||||
755 GGGCAAGGCGGGCGGCGGCGGCGGCTTGGTGGGTGACAGTCTTACTT 804
406 SerTrpSerPro...AlaSerIleProGlyLysHisTyrGlnAlaValGI 421
|||||
805 TCTGGAAGCCCAACGAGGACTTCCCGGGCGNAGCACTTACCAAGCTTGG 854
421 yLeu 422
|||
855 GGTT 858

seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AAH99362
seq_documentation_block:
ID AAH99362 standard; cDNA: 751 BP.
XX
AC AAH99362;
XX
DT 16-OCT-2001 (first entry)
XX
DE Human protein encoding cDNA sequence SEQ ID NO:197.
XX
KW Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
KW anti-inflammatory; antirheumatic; antiarthritic; immunosuppressive;
KW antibacterial; endocrine; cardiant; central nervous system; virucide;
KW anti-HIV; fungicide; antimutagen; cardiovascular; antianemic; anaemia;
KW antiaggregant; haemostatic; antiallergic; antidiabetic; eczema;
KW dermatological; antiallergic; antidiabetic; antidiabetic; infection;
KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
KW immunostimulant; gene therapy; antitense therapy; vaccine; inflammation;
KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
KW allergic rhinitis; diabetes; multiple sclerosis; depression;
KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
KW neurological disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200153455-A2.
XX
XX
PD 26-JUL-2001.
XX
XX
PF 22-DEC-2000; 2000WO-US35017.
XX
XX
PR 23-DEC-1999; 99US-0471275.
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
XX
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
XX
DR WPI; 2001-457603/49.
DR
DR
XX
XX
PT Isolated human polynucleotides encoding polypeptides, useful for the
PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
XX
PS Claim 1; Page 386; 1217pp; English.
XX
XX
CC AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
CC AAM25963. The proteins can have activities based on the tissues and
CC cells they are expressed in, such as: antinflammatory; antirheumatic;
CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;

CC central nervous system; virucide; anti-HIV; fungicide; antimitagen;
CC cardiovascular; antianemic; antiagregant; antiaagregant; antiaagregant;
CC antifungal; osteopathic; dermatological; antiallergic; antiasthmatic;
CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
CC encoding them can be used in gene therapy, antisense therapy and vaccine
CC production. The proteins and polynucleotides are useful for screening for
CC agonists or antagonists of a protein and for the treatment and diagnosis
CC of disorders associated with the activity of a protein e.g. inflammation,
CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
CC osteoporosis, severe combined immunodeficiency, eczema, allergic
CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
CC neurological disorders.
XX
SQ Sequence 751 BP; 133 A; 219 C; 255 G; 144 T; 0 other;

alignment_scores:
Quality: 964.00 Length: 195
Ratio: 5.446 Gaps: 1
Percent Similarity: 90.769 Percent Identity: 89.744

alignment_block:
US-09-768-877-2 x AAH99362 ..

Align seg 1/1 to: AAH99362 from: 1 to: 751

60 Proatrginglglglnvalylsglglleuileugllyaspctsrphphle 76
|||||
2 CCACGGGAAGGGCAGGTGAGCAGGGGCTGCTGGGATGCTGCTCT 51
76 ucysalacysalaalaleuainlyserserhisleuleuaspglinali 93
|||||
52 GTGTGCTGCGCGCGCTGCAGAGAGCAGCCTCTGAGCAGGTCA 101
93 leProProglglglnProserTrpalaaspglglutyrargllyserphe 109
|||||
102 TTCTCCCGGAGACCGGAGCTGGCCGACACAGAGTACCGGGCTCTTC 151
110 ThrcysarglletTrpGlnPheglYargTrpValgluValThrThrAspAs 126
|||||
152 ACCTGTCCATTTGGCAGTTGGAGCTGGGTGGAGGTGACACAGATGA 201
126 PargleuProCysleuAlaaglYargLeuCysPheSerArgCysglinarGg 143
|||||
202 CCGCTGCGCGCTGCTGCAGGGAGACTGTCTCCCGCTGCCAGAGGG 251
143 luaspvalPheTrpLeuProleuLeuGlulysValTyrAlaIysValHis 159
|||||
252 AGGATGTGTGCTGCTCCCTTACTGGAAGGTCTACGCCAGGTCCAT 301
160 GlyserTrpGlnHisLeuTrpAlaGlYglInValAlaAspAlaLeuValAs 176
|||||
302 GGGTCTCTACGACACCTGTGGCGGGCAGGTGGGGATGCCCTGTGCA 351
176 PleuThrGlyglYleuAlaGluarGTrpAsnLeuLysGlyValAlaGlys 193
|||||
352 CCTACCGCGCGCTGCGCAAAAGATGMACTGAAAGGCGCTACACAGAA 401
193 erglyglYglnglnAspArgProglYargTrpGlnHisArgThrCysArg 209
|||||
402 GCGAGGCGCAGACAGACAGCGCGCTGGAGCACAGGACTTGTGCG 451
210 GlnleuLeuHisleuLysaspGlnCysleuIleserCysValleuSe 226
|||||
452 CAGCTGCTCCACCTGAAGGACCAAGTGTGATCAGCTGCTGCTGCTAG 501
226 rProatrgAlaGlYalaArgGlulLeuGlYgluPheHisAlaPheIleValS 243
|||||

502 CCCACAGACAGGT 514
243 erAspleuArglglleuileuglglnglglAlaGlYglIn 254
|||
515GAGGCACGTGGCCACAGCATGAGGAGG 538

OM of: US-09-768-877-2 to: Issued_Patents_NA:* out_format : pfs

Date: Sep 17, 2002 6:34 AM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODEL=itame+pn.model -DEV=xlp
-/cg2_1/USPTO_POOL/US09768877/runat.16092002.152812.8369/app-query.fasta.1.739
-BA-issued_Patents_NA -OFMT=fasta -SUFFIX=ri -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000
-FCAPOP=6.000 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -YCAPOP=10.000 -YGAPEXT=0.500
-TRANS-human40.cdi -LIST=45 -DOCLIGN=200 -THR_SCORE=pcr
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09768877_ECGNL_1.98 -NCPD=6 -ICPD=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPHY -WAIT -THREADS=1
```

Search information block:

```
Query: US-09-768-877-2
Query length: 672
Database: Issued_Patents_NA:*
Database sequences: 383533
Database length: 122816752
Search time (sec): 84.830000
```

score_list:

```
Sequence Strd Orig ZScore EScore Len i Documentation
/cg2_6/ptodata/2/lna/6B.COMB.seq:US-09-422-869-3 + 3649.00 6014.72 0.0 2620
/cg2_6/ptodata/2/lna/6B.COMB.seq:US-09-422-869-13 + 3604.00 5940.88 0.0 2455
/cg2_6/ptodata/2/lna/6B.COMB.seq:US-09-422-869-5 + 3290.00 5420.79 3.0e-294 2297
/cg2_6/ptodata/2/lna/6B.COMB.seq:US-09-422-869-11 + 3229.50 5333.87 1.6e-289 2516
/cg2_6/ptodata/2/lna/6B.COMB.seq:US-09-422-869-9 + 3090.50 5090.34 7.6e-264 2204
/cg2_6/ptodata/2/lna/6B.COMB.seq:US-09-422-869-19 + 2963.00 4877.19 5.7e-264 2511
/cg2_6/ptodata/2/lna/6B.COMB.seq:US-09-422-869-7 + 2715.50 4469.43 2.9e-241 2001
/cg2_6/ptodata/2/lna/6B.COMB.seq:US-09-422-869-1 + 1989.00 3224.17 6.7e-172 49136
/cg2_6/ptodata/2/lna/6B.COMB.seq:US-08-835-099A-5 + 635.00 1017.26 5.6e-49 2109
/cg2_6/ptodata/2/lna/6B.COMB.seq:US-08-835-099A-6 + 635.00 1017.26 5.6e-49 2109
/cg2_6/ptodata/2/lna/6B.COMB.seq:US-09-157-349-6 + 633.00 1017.10 5.8e-49 2136
/cg2_6/ptodata/2/lna/6B.COMB.seq:US-09-157-349-9 + 633.00 1017.10 5.8e-49 2136
/cg2_6/ptodata/2/lna/6B.COMB.seq:US-09-422-869-15 + 513.50 822.07 4.2e-38 1267
/cg2_6/ptodata/2/lna/6B.COMB.seq:US-09-422-869-17 + 472.00 758.00 1.6e-34 864
/cg2_6/ptodata/2/lna/6B.COMB.seq:US-08-928-692-16 + 263.00 390.11 4.8e-14 4700
/cg2_6/ptodata/2/lna/6B.COMB.seq:US-08-658-136-2 + 164.00 195.46 0.0034 53526
/cg2_6/ptodata/2/lna/6B.COMB.seq:US-08-658-136-1 + 164.00 195.45 0.0034 53577
/cg2_6/ptodata/2/lna/6B.COMB.seq:US-08-943-731-5 + 151.00 186.15 0.0111 20084
/cg2_6/ptodata/2/lna/6B.COMB.seq:US-09-103-840A-1 + 147.50 203.08 0.0013 3257
/cg2_6/ptodata/2/lna/6B.COMB.seq:US-09-056-105-26 + 136.50 180.86 0.0219 4473
/cg2_6/ptodata/2/lna/6B.COMB.seq:US-08-229-515A-9 + 136.50 180.86 0.0219 4473
/cg2_6/ptodata/2/lna/6B.COMB.seq:US-08-645-865-9 + 136.50 180.71 0.0223 4530
/cg2_6/ptodata/2/lna/6B.COMB.seq:US-09-335-409-1 + 136.00 145.88 1.94 68750
/cg2_6/ptodata/2/lna/6B.COMB.seq:US-09-568-102-1 + 136.00 145.88 1.94 68750
/cg2_6/ptodata/2/lna/6B.COMB.seq:US-09-568-969-1 + 136.00 145.88 1.94 68750
/cg2_6/ptodata/2/lna/6B.COMB.seq:US-09-568-480-1 + 136.00 145.88 1.94 68750
/cg2_6/ptodata/2/lna/6B.COMB.seq:US-09-568-486-1 + 136.00 145.88 1.94 68750
/cg2_6/ptodata/2/lna/6B.COMB.seq:US-09-568-472-1 + 136.00 145.88 1.94 68750
/cg2_6/ptodata/2/lna/6B.COMB.seq:US-09-103-840A-1 + 134.50 91.38 2.1e+03 4411529
/cg2_6/ptodata/2/lna/6B.COMB.seq:US-08-476-866-20 + 132.50 164.07 0.1883 10079
/cg2_6/ptodata/2/lna/6B.COMB.seq:US-08-804-227C-7 + 132.50 145.55 2.03 44377
/cg2_6/ptodata/2/lna/6B.COMB.seq:US-08-804-198-1 + 132.50 145.55 2.03 44377
/cg2_6/ptodata/2/lna/6B.COMB.seq:US-08-188-582-31 + 132.00 175.11 0.0457 3901
/cg2_6/ptodata/2/lna/6B.COMB.seq:US-08-646-715-31 + 132.00 175.11 0.0457 3901
/cg2_6/ptodata/2/lna/6B.COMB.seq:US-09-150-460B-3 + 131.00 168.51 0.1066 6614
/cg2_6/ptodata/2/lna/6B.COMB.seq:US-09-443-501A-2 + 131.00 137.01 6.06 71989
/cg2_6/ptodata/2/lna/6B.COMB.seq:US-09-422-869-21 + 130.50 181.78 0.0194 1875
/cg2_6/ptodata/2/lna/6B.COMB.seq:US-08-306-691B-42 + 130.00 164.46 0.1792 7011
/cg2_6/ptodata/2/lna/6B.COMB.seq:US-09-103-840A-2 + 129.00 82.28 6.8e+03 4403765
```

```
/cg2_6/ptodata/2/lna/6B.COMB.seq:US-09-404-650-1 + 128.50 162.33 0.2356 681
/cg2_6/ptodata/2/lna/6B.COMB.seq:US-09-404-650-3 + 128.50 162.26 0.2378 685
/cg2_6/ptodata/2/lna/6B.COMB.seq:US-08-123-161A-13 + 128.00 174.60 0.0488 2
/cg2_6/ptodata/2/lna/6B.COMB.seq:US-08-483-278-13 + 128.00 174.60 0.0488 23
/cg2_6/ptodata/2/lna/6B.COMB.seq:US-08-068-945A-1 + 127.00 153.27 0.7531 11
```

seq_name: /cg2_6/ptodata/2/lna/6B.COMB.seq:US-09-422-869-3

```
seq_documentation_block:
Sequence 3, Application US/09422869
Patent No. 6235481
GENERAL INFORMATION:
APPLICANT: POLONSKY, KENNETH S.
APPLICANT: HORIKAWA, YUKIO
APPLICANT: ODA, NAOHISA
APPLICANT: COX, NANCY J.
APPLICANT: SREENAN, SEAMUS
APPLICANT: ZHOU, YUN-PING
APPLICANT: OTANI, KENICHI
APPLICANT: HANIS, CRAIG L.
APPLICANT: BELL, GRAEME I.
TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
FILE REFERENCE: ARCD:307
CURRENT APPLICATION NUMBER: US/09/422,869
CURRENT FILING DATE: 1999-10-21
EARLIER APPLICATION NUMBER: 60/134,175
EARLIER FILING DATE: 1999-05-13
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3
LENGTH: 2620
TYPE: DNA
ORGANISM: Human
US-09-422-869-3
```

```
alignment_scores:
Quality: 3649.00 Length: 672
Ratio: 5.430 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
```

alignment_block:
US-09-768-877-2 x US-09-422-869-3

Align seg 1/1 to: US-09-422-869-3 from: 1 to: 2620

```
1 MetAAGAlaGlyArgGlyAlaThrProAlaArgGluLeuPheArgAspAl 17
|||||
178 ATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 227
|||||
17 aaAlaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrPro 34
|||||
228 CGCCTTCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 277
|||||
34 euAlaGluPheArgGluAspIleThrTrpArgArgProGluIleCys 50
|||||
278 TGGCCCGAGTCCCGCGGAGATCATCGAGCGGCGGCGGCGGCGGAGATTGT 327
|||||
51 AlaThrProArgLeuPheProAspAspProArgGluGlnGlnVallysgl 67
|||||
328 GCCACACCCCGCGGCTGTTCCAGATGACCCAGGAGGAGGAGGAGGAGGCA 377
|||||
67 nglyLeuLeuGlyAspCysTrpPheLeuGlyAlaCysAlaAlaLeuGln 84
|||||
378 GGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 427
|||||
84 ySerTrpGHisLeuLeuAspGlnValIleProProGluGlnProSerTrp 100
|||||
428 AGGACGAGGACCTCTCGACAGGATCTCTCCGAGGAGCGGAGCGGCTGG 477
|||||
101 AlaAspGlnGluTrpArgGlySerPheThrCysArgGlyIleTrpGlnPhe 117
|||||
478 GCCGACGAGGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 527
|||||
```

117 yAGTrpValGluValThrThraspArgLeuProCysLeuAlaGlyA 134
528 ACGTGGTGGAGTGCACACAGATGACCGCTGCGTGCCTTGCAGGGA 577
134 rgLeuCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
578 GACTCTGTTTCTCCCGCTGCCAGAGGAGGATGTCTTGGCTCCCTTA 627
151 LeuGluLysValTyrAlaLysValHisGlySerTyrGluHisLeuTrpAl 167
628 CTGGAAAGGTCTACGCCAAGGTTCATGGGTCCCTACGAGCACCTGTGGGC 677
167 aGlyGlnValAlaAspAlaLeuValAspLeuThrGlyGlyLeuAlaGluA 184
678 CGGCGAGGTGGCGATGTCCTTGGTGCACCTGACCGCGCGCTTGGCAGAA 727
184 rgTrpAsnLeuLysGlyValAlaGlySerGlyGlyGlnGlnAspArgPro 200
728 GATGGAAACCTGAAGGCGGTAGCAGGAAGCGGAGGCCAGCAGACAGGCCA 777
201 GlyArgTrpGluHisArgThrCysArgGlnLeuLeuHisLeuLysAspGL 217
778 GCGCGCTGGGAGCACAGGACTTGTCCGAGAGTGTCTCACTTGAAGGACCA 827
217 nCysLeuIleSerCysValLeuSerProArgAlaGlyAlaArgGluL 234
828 GTCTCTGATCAGCTGCTGGTGCTCAGCCCCAGACGAGGTCCCGGGAGC 877
234 euGlyGluPheHisAlaPheIleValSerAspLeuArgGluLeuGlnGly 250
878 TGGGGAGTGTCCATGCTTCATTGTCTCGGACCTGCGGGAGCTCCAGGCT 927
251 GlnAlaGlyGlnCysIleLeuLeuArgIleGlnAsnProTrpGlyAr 267
928 CAGCGGGGCGATGTGATCTCTGCTGCGGATCCAGAAACCCCTGGGCGC 977
267 gArgCysTrpGlnGlyLeuTrpArgGluGlyGlyGluGlyTrpSerGlnV 284
978 GCGGTGCTGCGCAGGGGCTCTGGAGAGAGGGGGTGAAGGTGTGAGCCAGG 1027
284 alAspAlaValAlaSerGluLeuLeuSerGlnLeuGlnGluGlyGlu 300
1028 TAGATGCACGGTAGCATCTGAGTCTCTGCCAGCTCCAGGAAGGGGAG 1077
301 PheTrpValGluGluGluPheLeuArgGluPheAspGluLeuThrVa 317
1078 TTCTGGGTGGAGGAGGAGTCTCTCAGGGAGTTTTCACGAGCTCACCGT 1127
317 lGlyTrpProValThrGluAlaGlyHisLeuGlnSerLeuTyrThrGluA 334
1128 TGGCTACCGGTCACGGAGCGCGGCACCTGCAGAGAGCTCTACACAGAGA 1177
334 rgLeuLeuCysHisThrArgAlaLeuProGlyAlaIleTrpValLysGlyGln 350
1178 GCGTCTCTGCCATACGCGGGCGCTGCTGCGGGCTTGGTCAAGGGCCAG 1227
351 SerAlaGlyGlyCysArgAsnAsnSerGlyPheProSerAsnProLysph 367
1228 TCACGAGGAGGCTGCCGGAAACAACAGCGGGCTTTCCACAGCAACCCAAAT 1277
367 eTrpLeuArgValSerGluProSerGluValTyrIleAlaValLeuGlnA 384
1278 CTGGCTGCGGGTCTCAGAACCGAGTGAGGTGTACATTGCCGTCTCTGCAGA 1327
384 rgSerArgLeuHisAlaAlaAspTrpAlaGlyArgAlaArgAlaLeuVal 400
1328 GATCAGAGCTGCACGCGCGGACTTGGGACGCGCGGCGCGGCGGCGACTGGTG 1377
401 GlyAspSerHisThrSerTrpSerProAlaSerIleProGlyLysHisty 417
1378 GGTGACGTATACTCTGTGGAGCCCAAGCAGCATCTCCGGGCAAGCACTA 1427

417	rGlnAlaValcIyLeuHisLeuTrpLysValGluLysArgValAsnL	434
1428	CCAGGCTGTGGGTCGTGACCTCTGGAAGGTAGAGAAGCGCGGTC	1477
434	euProArgValLeuSerMetProProValAlaGlyThrAlaCysHisAla	450
1478	TGCCTAGGGTCCTGTCATGCCGCCCGTGGCTGSCACCGCGTGCCATGCA	1527
451	TyrAspArgGluValHisLeuArgCysGluLeuSerProGlyTyrTyrIle	467
1528	TACGACCGGGAGGTCCACCTGCGGTGTGAGCTCTCACCGGGCTACTACCT	1577
467	uAlaValProSerThrPheLeuLysAspAlaProGlyGluPheLeuAla	484
1578	GGCTGTCCCGACGACCTTCCTGAAGGACGCGCCAGGGGAGTTCCTGCTCC	1627
484	rgValPheSerThrGlyArgValSerLeuSerAlaIleArgAlaValAla	500
1628	GAGTCTTCTTACCGGGCGAGTCTCCCTTAGCGCCATCATGGCAGTGGCC	1677
501	LysAsnThrThrProGlyAlaAlaLeuProAlaGlyGluTrpGlyThrVal	517
1678	AAGAACACCAACCCCGGGGCGAGCCCTGCTGCCGGGGAGTGGGGACCGT	1727
517	lGlnLeuArgGlySerTrpArgValGlyGlnThrAlaGlyGlySerArgA	534
1728	GCAGCTACGGGGTCTTGGAGAGTCGGCCAGACGCGGGGGGCGACGAGCA	1777
534	snPheAlaSerTyrProThrAsnProCysPheProPheSerValProGlu	550
1778	ACTTGTGCTCATACCCACCAACCCCTGCTTCCCTTCCTCGTCCCGCAG	1827
551	GlyProGlyProArgCysValArgIleThrLeuHisGlnHisCysArgPr	567
1828	GGCCCTGGCCCCCGCTGGTCCGCATCACTTCATCAGCACTGCGGGCC	1877
567	oSerAspThrGluPheHisProIleGlyPheHisIlePheGlnValProG	584
1878	CAGTGACACCGAGTTCCACCCATCCGCTTCCATATCTTCCAGGTCCCA	1927
584	luGlyArgSerGlnAspAlaProProLeuLeuGlnGluProLeu	600
1928	AGGCTGGAAGGAGCCAGGACGACCCCACTGCTGCTGACGAGGCGGTG	1977
601	LeuSerCysValProHisArgTyrAlaGlnGluValSerArgLeuCysIle	617
1978	CTGAGCTGGTGCCACATCGCTAGGCCACGAGGAGTGACCGGCTCTGCCT	2027
617	uLeuProAlaGlyThrTyrLysValValProSerThrTyrLeuProAspT	634
2028	CCTGCTCGCAGGCACCTACAAGGTTGTGCCCTCCACCTACCTGCCGACA	2077
634	hrIleGlyAlaPheThrValThrIleAlaThrArgIleAspArgProSer	650
2078	CAGAGGGGGCCCTTCACAGTGAACCTCGCAACACAGATTGACAGGCCATCC	2127
651	IleHisSerGlnGluMetLeuGlyGlnPheLeuGlnGluValSerValIle	667
2128	ATTACAGCCAGGAGTGTCTGGGCCAGTCTCTCCAAGAGGTCTCCGTCT	2177
667	tAlaValMetLysThr	672
1778	GGCAGTGTGATAAAACC	2193

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-422-869-13

seq_documentation_block:

; Sequence 13, Application US/09422869

; Patent No. 6235481

GENERAL INFORMATION:

APPLICANT: POLONSKY, KENNETH S.

; APPLICANT: HORIKAWA, YUKIO

; APPLICANT: ODA, NAOHISA

APPLICANT: COX, NANCY J.
APPLICANT: SREENAN, SEAMUS
APPLICANT: ZHOU, YUN-PING
APPLICANT: OTANI, KENICHI
APPLICANT: HANIS, CRAIG L.
APPLICANT: BELL, GRAEME I.
TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
FILE REFERENCE: ARCD:307
CURRENT APPLICATION NUMBER: US/09/422,869
EARLIER FILING DATE: 1999-10-21
EARLIER APPLICATION NUMBER: 60/134,175
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 13
LENGTH: 2455
TYPE: DNA
ORGANISM: Human
US-09-422-869-13

alignment_scores:
Quality: 3604.00 Length: 673
Ratio: 5.395 Gaps: 2
Percent Similarity: 99.257 Percent Identity: 99.257

alignment_block:

US-09-768-877-2 x US-09-422-869-13 ..

Align seg 1/1 to: US-09-422-869-13 from: 1 to: 2455

1 MetArgLaGlyArgGlyAlaThrProAlaArgGluLeuPheArgAspAl 17
|||||
42 ATGGGGCGGGCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 91
17 AlaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrProL 34
|||||
92 CGCCTTCGCGCGCGGAGCTCCGCTCTTGAGCTGTGCTACGCGCGC 141
34 euAlaGlnPheArgGluAspLeuThrThrParArgProGlnGluIleCys 50
|||||
142 TGGCCCGCTCCGCGGAGACATCAGTGAGGCGCGCGCGCGCGGAGATTGT 191
51 AlaThrProArgLeuPheProAspAspProArgGluGlnValLysGln 67
|||||
192 GCCACACCGCGCGCTGTTCCAGATGACCCAGGAGGAGGAGGAGGAGCA 241
67 ngLyleuLeuGlyAspCysThrPheLeuCysAlaCysAlaAlaLeuGln 84
|||||
242 GGGCTGTGCTGGGGGATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 291
84 ySerArgHisLeuLeuAspGlnValIleProProGlnGlnProSerTrp 100
|||||
292 AGAGCAGGCGACCTCTGACAGCATCTCTCGGAGACCGAGAGTGG 341
101 AlaAspGlnGluThrArgGlySerPheThrCysArgIleTrpIlePheG 117
|||||
342 GCCGACAGAGATACCGGGCTCTTCACTGTCTGCAATTTGGCAGTTGG 391
117 yArgTrpValGluValThrThrAspAspArgLeuProCysLeuAlaGly 134
|||||
392 AGCGTGGTGGAGGTACACAGATACCGCTGCGGTGCTTGGCAGGGA 441
134 rGluCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
|||||
442 GACTGTGTTCTCCGCTCCAGAGGAGAGATGTGTTGGCCCTTCA 491
151 LeuGluLysValThrAlaLysValHisGlySerTrpGluHisLeuTrp 167
|||||
492 CTGGGA.....AAGGCTCATGTGGTCTACGAGACACTGTGGG 529
167 LaGlyGlnValAlaAspAlaLeuValAspLeuThrGlyLeuAlaGlu 183
|||||

530 CCGGGCAGGTGGCGGATGCCCTGTGTGACCTGACCGCGCGCGCTGGCAGNA 579
184 ArgTrpAsnLeuLysGlyValAlaIleSerGlyGlyGlnGlnAspArgPr 200
|||||
580 AGATGAACTGAGGGCGGTACAGAGAACCGGAGGCGCACAGACAGCGCC 629
200 oGlyArgTrpGlnHisArgThrCysArgGlnLeuHisLeuLysAspG 217
|||||
630 AGCGCGCTGGGAGACACAGACTGTTCGAGAGCTGCTCCACCTGAGAGGAC 679
217 lncysLeuIleSerCysCysValLeuSerProArgAlaGlyAlaArgGlu 233
680 AGTGTCTATCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 729
234 LeuGlyGluPheHisAlaPheIleValSerAspLeuArgGluLeuGln 250
|||||
730 CTGGGGAGTTCCATGCTTCTATGCTTCGAGACCTGCGGAGCTCCAGAG 779
250 yGlnAlaGlylncysIleLeuLeuLeuArgIleGlnAsnProTrpGly 267
780 TCAGCGGGCCAGTGCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 829
267 rArgCysTrpGlnGlyLeuThrParArgGluGlyGlyGlyTrpSerGln 283
830 GCGGTGCTGCGAGGGGCTGTGAGAGAGGGGGTGAAGGGTGGAGCCAG 879
284 ValAspAlaAlaValAlaSerGluLeuLeuSerGlnLeuGlnGlyGly 300
880 GTATATGACGCGGTAGCATCTGACCTCTGCTGCTGCTGCTGCTGCTG 929
300 uPheTrpValGluGluGluGluPheLeuArgGluPheAspGluLeuThr 317
930 GTCTGTGGTGGAGAGAGAGAGGTTCTCTCAGGAGTTTACAGACTCACCG 979
317 aLgLyTrpProValThrGluAlaGlyHisLeuGlnSerLeuTrpThrGlu 333
980 TTGGCTACCGCGTCCAGAGGCGCGCCACCTGCAGAGCTCTACACAGAG 1029
334 ArgLeuLeuCysHisThrArgAlaLeuProGlyAlaTrpValLysGly 350
1030 AGGTGCTCTGCTCCATACCGCGCGCTGCTGCTGCTGCTGCTGCTGCT 1079
350 nSerAlaGlylncysArgAsnAsnSerGlyPheProSerAsnProLysP 367
1080 GTGAGCAGAGAGCGCGCGGAAACACACGCGCTTCCAGCAACCCCAAT 1129
367 heTrpLeuArgValSerGluProSerGluValTrpIleAlaValLeuGln 383
1130 TCTGGCTGCGGGTCTCAGAACCGAGTGAGGTGTCATTTGCGTCTGAG 1179
384 ArgSerArgLeuHisAlaAlaAspTrpAlaGlyPheArgAlaArgAlaLeu 400
1180 AGATTCAGGCTGACGCGGCGGACTGTGGGAGGCGCGCGCGGACACTGT 1229
400 lGlyAspSerHisThrSerTrpSerProAlaSerIleProGlyLysHis 417
1230 GGTGACAGTACATCTTCTGTGAGAGCCAGGAGACATCTCCGGGAAGACT 1279
417 yGlnAlaValGlyLeuHisLeuTrpLysValGlnLysArgValAsn 433
1280 ACCAGGCTGTGGGTGTGACCTCTGGAAGGTAAAGAAAGCGCGGCTCAAT 1329
434 LeuProArgValLeuSerMetProProValAlaGlyThrAlaCysHisAl 450
1330 CTGCTAGGGTCTCTTCATGTGCCCCGCTGCTGCTGCTGCTGCTGCTG 1379
450 aTyraAspArgGluValHisLeuArgCysGlnLeuSerProGlyTrpTyrl 467
1380 ATACGACCGGAGGTCCACCTCGTTGTAGCTTCCACCGGGGTACTACAC 1429
467 euAlaValProSerThrPheLeuLysAspAlaProGlyGluPheLeu 483
1430 TGGCTGTCCCGACACCTTCTGGAAGAGCGCGCAGGAGTTCTGCTC 1479

```

484 ArgValPheSerThrGlyArgValSerLeuSerAlaIleArgAlaValAl 500
|||||
1480 CGAGTCTTCTACCGGGGAGTCTCCCTAGCGCATCAGGCGAGTGGC 1529
|||||
500 alyAsnThrThrProGlyValAlaLeuProAlaGlyGluThrGlyThrV 517
|||||
1530 CAAGAACACACCCCGGGGAGCCCTGCTCGGGGAGTGGGGACCG 1579
|||||
517 aGlnLeuArgGlySerThrArgValGlyGlnThrAlaGlyGlySerArg 533
|||||
1580 TGCAGCTACGGGTTCTTGGAGAGTGGCCAGACGGGGGGGACAGG 1629
|||||
534 AsnPheAlaSerTyrProThrAsnProCysPheProPheSerValProG 550
|||||
1630 AACTTGGCTCATACCCACCAACCCCTGCTTCCCTTCTCGGTCCCGA 1679
|||||
550 uGlyProGlyProArgCysValArgIleThrLeuHisGlnHisCysArgP 567
|||||
1680 GGGCCTGGCCCGCTGGTCCGCATCCTGTCATCAGCACTGCCGCG 1729
|||||
567 roSerAspThrGluPheHisProIleGlyPheHisIlePheGlnValPro 583
|||||
1730 CCAGTGACACCGAGTCCACCCATCGGCTTCCATATCTTCAGGTCCA 1779
|||||
584 GluGlyGlyArgSerGlnAspAlaProProLeuLeuGlnGluProle 600
|||||
1780 GAGGTGGAGGAGCCAGGACCCACCCCTGCTGCTGAGGAGCCGCT 1829
|||||
600 uLeuSerCysValProHisArgTyrAlaGlnGluValSerArgLeuCysL 617
|||||
1830 GCTGAGTGGTGCACATCGCTACGCCAGGAGGTGAGCGGCTCTGCC 1879
|||||
617 euLeuProIaGlyThrTyrLysValValProSerThrTyrLeuProAsp 633
|||||
1880 TCCTGCTCAGGACCCCTACAAGGTGTGCCCTCCACCTACCTGCCGAC 1929
|||||
634 ThrGluGlyAlaPheThrValThrIleAlaThrArgIleAspArgProSe 650
|||||
1930 ACAGAGGGGGCTTACAGTGACCATCGCACCCAGGAGTTCAGGCGCATC 1979
|||||
650 rIleHisSerGlnGluMetLeuGlyGlnPheLeuGlnGluValSerValM 667
|||||
1980 CATTCACAGCCAGGAGATGCTGGCCAGTTCCTCCAGAGGTCTCCGTCA 2029
|||||
667 etAlaValMetLysThr 672
2030 TGGCAGTGATGAACCC 2046

```

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-422-869-5'

seq_documentation_block:

; Sequence 5, Application US/09422869

; Patent No. 6235481

; GENERAL INFORMATION:

; APPLICANT: POLONSKY, KENNETH S.

; APPLICANT: HORIKAWA, YUKIO

; APPLICANT: ODA, NAOHISA

; APPLICANT: COX, NANCY J.

; APPLICANT: SREENAN, SEAMUS

; APPLICANT: ZHOU, YUN-PING

; APPLICANT: OTANI, KENICHI

; APPLICANT: HANIS, CRAIG L.

; APPLICANT: BELL, GRAEME I.

; TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES

; FILE REFERENCE: ARCD:307

; CURRENT APPLICATION NUMBER: US/09/422,869

; CURRENT FILING DATE: 1999-10-21

; EARLIER APPLICATION NUMBER: 60/134,175

; EARLIER FILING DATE: 1999-05-13

; NUMBER OF SEQ ID NOS: 30

; SOFTWARE: PatentIn ver. 2.0

; SEQ ID NO 5

```

; LENGTH: 2297
; TYPE: DNA
; ORGANISM: Human
US-09-422-869-5

alignment_scores:
  Quality: 3290.00      Length: 672
  Ratio: 5.350          Gaps: 1
Percent Similarity: 91.518 Percent Identity: 91.518

alignment_block:
US-09-768-877-2 x US-09-422-869-5

```

Align seg 1/1 to: US-09-422-869-5 from: 1 to: 2297

```

1 MetArgAlaGlyArgGlyAlaThrProAlaArgGluLeuPheArgAspAl 17
|||||
42 ATGCGGGGCGGGCGGGCGGACGCGCGGGAGGAGCTGTTCGGGAGCGC 91
|||||
17 aAlaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrProL 34
|||||
92 CGCCTTCCCGCGCGGACTCTCTGCTCTTCTGCGACTTGTCTACGCCGC 141
|||||
34 euAlaGlnPheArgGluAspIleThrTrpArgArgProGlnGluIleCys 50
|||||
142 TGGCCCACTTCGCGGAGGACATCAGTGGAGGCGGCCCGGAGATTGT 191
|||||
51 AlaThrProArgLeuPheProAspProArgGluGlyGlnValLysG 67
|||||
192 GCCACACCCCGGCTGTTCCAGATGACCCACGGGAGGCGAGGTGAGCA 241
|||||
67 nGlyLeuLeuGlyAspCysTyrPheLeuCysAlaCysAlaLeuGlnL 84
|||||
242 GGGGTGCTGGGGATTGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTG 291
|||||
84 ySerArgHisLeuLeuAspGlnValIleProProGlyGlnProSerTrp 100
|||||
292 AGAGCAGCAGCTCTTGGACCAAGGTGCTTCTCGGGGACAGCCGAGCTGG 341
|||||
101 AlaAspGlnGluTyrArgGlySerPheThrCysArgIleTrpGlnPheG 117
|||||
342 GCCGACAGGAGTACCGGGGCTCTTCCCTTCCCTTCCCTTCCCTTCC 391
|||||
117 yArgTrpValGluValThrThrAspAspArgLeuProCysLeuAlaGly 134
|||||
392 ACCTGGGTGGAGGTGACACAGATGACCGCTGCTGCTGCTGCTGCTG 441
|||||
134 rgLeuCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLe 150
|||||
442 GACTCTGTTTCTCGCTGCCAGAGGAGGAGTGTCTGCTGCTGCTGCT 491
|||||
151 LeuGluLysValTyrAlaLysValHisGlySerTyrGluHisLeuTrpAl 167
|||||
492 CTGGAAGAGTCTACGCCAAGGTCCATGGGTCTTACGAGCACCTGTGGG 541
|||||
167 aGlyGlnValAlaAspAlaLeuValAspLeuThrGlyGlyLeuAlaGlu 184
|||||
542 CGGCAGGTGGCGGATGCCCTGGTGGACCTGACCGGGGCGCTGGCAGAA 591
|||||
184 rgTrpAsnLeuLysGlyValAlaGlySerGlyGlyGlnGlnAspArgPro 200
|||||
592 GATGGAACCTGAAGGGCGGTACAGAGGAGGAGGAGGAGGAGGAGGAG 641
|||||
201 GlyArgTrpGluHisArgThrCysArgGlnLeuLeuHisLeuLysAspG 217
|||||
642 GGCGGTGGGAGCACAGGACTTGTGGGAGCTGTCCACCTGAAGGACCA 691
|||||
217 nCysLeuIleSerCysValLeuSerProArgAlaGlyAlaArgGluL 234
|||||
692 GTGCTGATCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 741
|||||
234 euGlyGluPheHisAlaPheIleValSerLeuArgGluLeuGlnGly 250

```

```

|||||
742 TGGGGGAGTTCATGCTTCATGCTGCGACTGCGGAGACTCCAGGT 791
251 GlnAlaGlyGlnCysIleLeuLeuLeuArgIleGlnAsnProTrpGlyIar 267
792 CAGCGGGGAGTGCATCTGCTGCTGCGGATCCAGAACCCCTGGGGCCG 841
267 GATGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 284
842 GCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 891
284 AlaAlaAlaAlaAlaAlaSerGluLeuLeuSerGlnLeuGlnGlyGlu 300
892 TAAATGACGCGGTAGCATCTGACCTCTGCTGCTGCTGCTGCTGCTG 941
301 PheTrpValGluGluGluGluPheLeuArgGluPheAspGluLeuThrVa 317
942 TTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 991
317 IglYTrpProValThrGluAlaGlyHisLeuGlnSerLeuTrpThrGlu 334
992 TGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1041
334 IglLeuLeuGlyHisThrArgAlaLeuProGlyAlaTrpValGlyGln 350
1042 GGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1091
351 SerAlaGlyGlyCysArgAsnAsnSerGlyPheProSerAsnProLysPh 367
1092 TCAGCGAGGAGGCTGCGGAGAACACACGCTTCCAGCAACCCCAAT 1141
367 eTrpLeuArgValSerGluProSerGluValTrpIleAlaValLeuGln 384
1142 CTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1191
384 rGSerArgLeuHisAlaAlaAspTrpAlaGlyArgAlaArgAlaLeuVal 400
1192 GATCCAGGCTGCGACGCGGCGACTGGCGCGCGCGCGCGCTGGTG 1241
401 GlyAspSerHisThrSerTrpSerProAlaSerIleProGlyLysHisTy 417
1242 GGTGACAGTCAATCTTCTGAGGAGCCAGCAGCATCCGCGGCAACAT 1291
417 rGlnAlaValGlyLeuHisLeuTrpLysValGluLysArgArgValAsnL 434
1292 CCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1341
434 eUrProArgValLeuSerMetProProValAlaGlyThrAlaCysHisAla 450
1342 TGCCAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1391
451 TyrAspArgGluValHisLeuArgCysGluLeuSerProGlyTrpTyrLe 467
1392 TACGACGCGGAGTCCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1441
467 uAlaValProSerThrPheLeuLysAspAlaProGlyGluPheLeuVal 484
1442 GCGTGTCCCGCAGCACTCTCTGAAAGAGAGCGCCGAGGAGTCTGCTCC 1491
484 rValAlaPheSerThrGlyArgValSerLeuSerAlaIleArgAlaValAla 500
1492 GAGTCTTCTTACCGCGGAGTCTCCCTTAG..... 1522
501 LysAsnThrThrProGlyAlaAlaLeuProAlaGlyLutTrpGlyThrVa 517
1522 ..... 1522
517 IglInLeuArgGlySerTrpArgValGlyGlnThrAlaGlyGlySerArg 534
1522 ..... 1522
534 snPheAlaSerTyrProThrAsnProCysPheProPheSerValProGlu 550

```

```

1522 ..... 1522
551 GlyProGlyProArgCysValArgIleThrLeuHisGlnHisCysArgP 567
1523 GCGCTGTGCGCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1572
567 oSerAspThrGluPheHisProIleGlyPheHisIlePheGlnValProG 584
1573 CAGTGCACACGAGTCCACCCCATCGGCTTCCATCTTCACAGTCCAG 1622
584 LuGlyGlyArgSerGlnAspAlaProProLeuLeuLeuGlnGluProLeu 600
1623 AGGCTGAAAGAGCCAGAGAGCCAGCCCTGCTGCTGCTGCTGCTGCTG 1672
601 LeuSerCysValProHisArgTyrAlaGlnGluValSerArgLeuGlyS 617
1673 CTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1722
617 uLeuProAlaGlyThrTrpLysValValProSerThrTrpLeuProAsp 634
1723 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1772
634 hGluGlyAlaPheThrValThrIleAlaThrArgIleAspArgProSer 650
1773 CAGAGGGGCTTGCACAGTGCATGCAACAGATTCGACAGGATTCG 1822
651 IleHisSerGlnGluMetLeuGlnPheLeuGlnGluValSerValMe 667
1823 ATTCAAGCCAGCAGAGATGCTGGGCCAGTTCCTCCAGAGAGTCTCC 1872
667 tAlaValMetLysThr 672
1873 GCGAGTGAATGAAC 1888
seq_name: /cgn2_6/ptodata/2/1na/6B_COMB.seq:US-09-422-869-11
seq_documentation_block:
: Sequence 11, Application US/09422869
: Patent No. 6235481
: GENERAL INFORMATION:
: APPLICANT: POLONSKY, KENNETH S.
: APPLICANT: HORIKAWA, YUKIO
: APPLICANT: ODA, NAOHISA
: APPLICANT: COX, NANCY J.
: APPLICANT: SREENAN, SEAMUS
: APPLICANT: ZHOU, YUN-PING
: APPLICANT: HANIS, KENICHI
: APPLICANT: HANIS, CRAIG L.
: TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
: FILE REFERENCE: ARCD:307
: CURRENT FILING DATE: 1999-10-21
: EARLIER APPLICATION NUMBER: 60/134,175
: EARLIER FILING DATE: 1999-05-13
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 11
: LENGTH: 2516
: TYPE: DNA
: ORGANISM: Human
US-09-422-869-11

```

alignment_scores:

```

Quality: 3239.50 Length: 745
Ratio: 5.267 Gaps: 2
Percent Similarity: 82.550 Percent Identity: 82.416

```

alignment_block:

```

US-09-768-877-2 x US-09-422-869-11 ..

```

```

Align seg 1/1 to: US-09-422-869-11 from: 1 to: 2516

```

1 MetArgAlaGlyArgGlyAlaThrProAlaArgGluLeuPheArgAspAl 17
 42 ATCGGGCGGGCGGGCGGCGAGCGCGGAGGAGCTGTTCCGGGACGC 91
 17 aAlaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrProL 34
 92 CGCCTTCCCGCGCGGAGCTCTCGCTCTCTCGCACTTGTCTACGCCGC 141
 34 euAlaGlnPheArgGluAspIleThrTrpArgArgProGlnGluLeuLeu 50
 142 TGGCCCAAGTTCGCGAGGACATCATCGTGGAGCGGCGCCCGAGGATTTGT 191
 51 AlaThrProArgLeuPheProAspAspProArgGluGlyGlnValysG1 67
 192 GGCACACCCCGGCTGTTTCCAGATGACCCACCGGAAGGCGCGTGAAGCA 241
 67 nGlyLeuLeuGlyAspCysTrpPheLeuLeuCysAlaCysAlaAlaLeuGlnL 84
 242 GGGGCTGCTGGGGGATGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 291
 84 ysSerArgHisLeuLeuAspGlnValIleProProGlyGlnProSerTrp 100
 292 AGACGAGGACCTCTCGGACGAGTCACTCTCCCGGACGAGCGAGCTGG 341
 101 AlaAspGlnGluTyrArgGlySerPheThrCysArgIleTrpGlnPheG1 117
 342 GCGCACGAGGAGTACCGGGCTCTTCACTGCTGCTGCTGCTGCTGCTGCTG 391
 117 yArgTrpValGluValThrThrAspAspArgLeuProCysLeuAlaGlyA 134
 392 ACCTGGGTGGAGGTGACACAGATGACCGCTGCTGCTGCTGCTGCTGCTG 441
 134 rGLeuCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
 442 GACTCTGTTCTCCCGCTCCAGAGGAGGATGTTCTGGCTCCCTTCA 491
 151 LeuGluLysValTyrAlaLysValHisGlySerTyrGluHisLeuTrpAl 167
 492 CTGGAAGAGTCTAGCCCAAGGTCTTCACTGCTGCTGCTGCTGCTGCTG 541
 167 aGlyGlnValAlaAspAlaLeuValAspLeuThrGlyLeuAlaGluA 184
 542 CGGGCAGGTGGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 591
 184 rGTrpAsnLeuLysGlyValAlaGlySerGlyGlnGlnAspArgPro 200
 592 GATGGAACCTGAAGGCGTAGCAGAGAGGAGGAGCGGACGACAGGCGCA 641
 201 GlyArgTrpGluHisArgThrCysArgGlnLeuLeuHisLeuLysAspG1 217
 642 GCGCGCTGGGAGCACAGGACTTGTGGCAGCTGCTCCACCTGAAGGACCA 691
 217 nCysLeuIleSerCysValLeuSerProArgAlaGlyAlaArgGluL 234
 692 GTGTCTGATCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 741
 234 euGlyGluPheHisAlaPheIleValSerAspLeuArgGluLeuGlnGly 250
 742 TGGGGAGTTCATGCCITTCATGCTCGGACCTCGGGAGCTCCAGGGT 791
 251 GlnAlaGlyGlnCysIleLeuLeuLeuArgIleGlnAsnProTrpGlyAr 267
 792 CAGGGGGCGCAGTGCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 841
 267 gArgCysTrpGlnGlyLeuTrpArgGluGlyGlyGlyGlyGlyGlyGly 284
 842 GCGGTGCTGGCAGGGGCTCTGGAGAGAGGGGGTGAAGGGTGGAGCCAGG 891
 284 aLaspAlaAlaValAlaSerGluLeuLeuSerGlnLeuGlnGluGlyGlu 300
 892 TAGATGCAGCGGTAGCATCTGAGCTCTCTCCAGCTCCAGGAGGGGAG 941

301 PheTrpValGluGluGluGluPheLeuArgGluPheAspGluLeuThrVa 317
 942 TTCTGGGTGGAGGAGGAGGAGTCTCTAGGAGGATTTGAGAGCTCACCGT 991
 317 lGlyTyrProValThrGluAlaGlyHisLeuGlnSerLeuTyrThrGluA 334
 992 TGGCTACCCGGTACGAGGAGCGGCGCACCTGCAGAGCCTCTACACAGAGA 1041
 334 rGLeuLeuCysHisThrArgAlaLeuProGlyAlaTrpValysGlyGln 350
 1042 GCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1091
 351 SerAlaGlyGlyCysArgAsnAsnSerGlyPheProSerAsnProLysPh 367
 1092 TCAGCAGGAGGCTGCGGAGACACAGCGGCTTTCACAGACACCCCAANT 1141
 367 eTrpLeuArgValSerGluProSerGluValTyrIleAlaValLeuGlnA 384
 1142 CTGGCTGGGGTCTCAAAACCGAGTGAGGTGTACATTGCGGTCTGCAGA 1191
 384 rGSerArgLeuHisAlaAlaAspTrpAlaGlyArgAlaArgAlaLeuVal 400
 1192 GATCCAGGCTGCACGCGGCGGACTGGGCGAGCGCGGCGGCGGCGGCTG 1241
 401 GlyAspSerHisThrSerTrpSerProAlaSerIleProGlyLysHisTy 417
 1242 GGTGACAGTCACTCTCTGGAGCGGCGGAGCATCCCGGCGGAGCAGCTA 1291
 417 rGlnAlaValGlyLeuHisLeuTrpLys..... 426
 1292 CCAGCTGTGGGTCTGCACCTCTGGAAGGCTGTGACACTTGGCACCCACAC 1341
 426 426
 1342 TGTTCCTGCTCCCTTCATGGATGTGGCCACATGATGTTCCTTCTCTT 1391
 426 426
 1392 GCAAAAGAAGTGTCTGGAAGGCCCTCTCCAGCAGCCCGGAGTTGGCT 1441
 426 426
 1442 GGGCCACGCTGCTTGTGGGCCAGCTTACAAGGAGGACTTGCAGGCTCG 1491
 427Val 427
 1492 TGTCTGGCAGAGTACTGCGCCGAGGCGCAAGTGAAGCCCGGGATTTGTA 1541
 428 GluLysArgArgValAsnLeuProArgValLeuSerMetProProValAl 444
 1542 GAGAAGCGCGGGTCAATCTGCTAGGCTCTGTCCATGCCCCCGGTGGC 1591
 444 aGlyThrAlaCysHisAlaTyrAspArgGluValHisLeuArgCysGluL 461
 1592 TGGCACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1641
 461 euSerProGlyTyrTrpLeuAlaValProSerThrPheLeuLysAspAla 477
 1642 TCTCAGCGGGTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1691
 478 ProGlyGluPheLeuLeuArgValPheSerThrGlyArgValSerLeuSe 494
 1692 CCAGGGGAGTTCTGCTCCGAGTCTCTCTACCGGCGGAGTCTCCCTTAG 1740
 494 rAlaIleArgAlaValAlaLysAsnThrThrProGlyAlaAlaLeuProA 511
 1741 1741
 511 lAGlyGluTrpGlyThrValGlnLeuArgGlySerTrpArgValGlyGln 527
 1741 1741
 528 ThrAlaGlyGlySerArgAsnPheAlaSerTyrProThrAsnProCysPh 544

```

1741 ..... 1741
544 eproheservalProgluGlyProargCysValargIlethrl 561
1742 ..... GGCCTGGCCGCCGCTGCTGCGCATCATCTC 1772
561 euHISglHIScysArpProserAspThrGluPheHisProIleGlyPhe 577
1773 TGCAATCAGACAGTCCGCCGACAGTACAGAGATTCCACCCATCGGCTTC 1822
578 HISIlePheGlnValProgluGlyGlyArgSerGlnAspAlaProProle 594
1823 CATACTTCCAGGTCCAGAGGGTGGAGAGCCAGCGACGCCACCCCTACT 1872
594 uLeuGlnGluProLeuLeuSerCysValProHisArgTyrAlaGln 611
1873 GCTGCTGAGAGACCGCTGCTGAGCTGCTGACATCGTACAGTCCAGG 1922
611 luValSerArgLeuGlyLeuLeuProAlaGlyThrTyrLeuValValPro 627
1923 AGGTAGCGCGGCTGCTGCTGCTGCTGACAGGACCTACAGTGTGTGCC 1972
628 SerThrTyrLeuProAspThrGluGlyAlaPheThrValThrIleAlaTh 644
1973 TCACACTACCTGCGCGACACAGAGGGGCTTCACAGTACCATCGCMAAC 2022
644 fArgIleAspArgProSerIleHisSerGlnGluMetLeuGlnIlePhe 661
2023 CAGATTTCAGAGGCATCATCTTCACAGCCAGAGATGCTGGGCCAGTTC 2072
661 euGlnGluValSerValMetAlaValMetLysThr 672
2073 TCCAAGAGGTCTCCGTCAATGAGATGATAAACC 2107

```

seq_name: /cgn2_6/ptodata/2/lna/bb_COMB.seq:US-09-422-869-9

```

seq_documentation_block:
; Sequence 9, Application US/09422869
; Patent No. 6235481
; GENERAL INFORMATION:
; APPLICANT: POLONSKY, KENNETH S.
; APPLICANT: HORIKAWA, YUKIO
; APPLICANT: ODA, NAOHISA
; APPLICANT: COX, NANCY J.
; APPLICANT: SREENAN, SEAMUS
; APPLICANT: ZHOU, YUN-PING
; APPLICANT: OTANI, KENICHI
; APPLICANT: HANIS, CRAIG L.
; APPLICANT: BELL, GRAEME I.
; TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
; FILE REFERENCE: ARCD:307
; CURRENT APPLICATION NUMBER: US/09/422,869
; EARLIER APPLICATION NUMBER: 60/134,175
; EARLIER FILING DATE: 1999-05-13
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 9
; LENGTH: 2204
; TYPE: DNA
; ORGANISM: Human
US-09-422-869-9

```

```

alignment_scores:
Quality: 3090.50      Length: 672
Ratio: 5.292          Gaps: 1
Percent Similarity: 86.905  Percent Identity: 86.905

```

alignment_block:

US-09-768-877-2 x US-09-422-869-9

Align seq 1/1 to: US-09-422-869-9 from: 1 to: 2204

```

1 MetArgAlaGlyArgGlyAlaThrProAlaArgGluLeuPheArgAspAl 17
42 AVGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 91
17 aAlaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrPro 34
92 GCGCTTCCCGCCCGGAGCTCTGCTCTTCTGCGACTGTGTACGGCCG 141
34 euAlaGlnPheArgGluAspIleThrTrpArgArgProGlnGluIleCys 50
142 TGCCCGAGTTCGGGAGAGATCACTGAGAGCGGGCCCGCAGAGATTTGT 191
51 AlaThrProArgLeuPheProAspAspProArgGluGlnValLysGln 67
192 GCCACACCCCGGCTGTTCCAGATGACCCACGGGAGAGGCGAGTGAAGA 241
67 nGlyLeuGlnGlyAspCysTrpPheLeuCysAlaCysAlaIleGlnIle 84
242 GGGGCTGCTGGGGGATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCA 291
84 ySSerArgHisLeuLeuAspGlnValIleProProGlnGlnProSerTrp 100
292 AGAGCAGGACACCTCTGAGACAGGTCAATTCCTCGGGACAGCGAGCTGG 341
101 AlaAspGlnGluTyrArgGlySerPheThrCysArgIleTrpGlnPheG 117
342 GCCGACCAAGAGTACCGGGGCTCTTCACCTGTCGATTTGGCAGTTTGG 391
117 yArgTrpValGluValThrThrAspAspArgLeuProCysLeuAlaGly 134
392 ACGGTGGGTGAGGTGACACAGATGACCGCTCCCTGCTTCAGAGGA 441
134 rGlyLeuCysPheSerArgCysGlnArgGluAspValPheTrpLeuPro 150
442 GACTCTGTTTCTCCCGCTCCAGAGAGAGATGTGCTGCTGCTCCCTTA 491
151 LeuGluLysValTyrAlaLysValHisGlySerTyrGlnHisLeuTrpAl 167
492 CTGAAAGGTCTAGCCAAAGTTCATGGGCTCTACAGACACCTGTGGGC 541
167 aGlyGlnValAlaAspAlaLeuValAspLeuThrGlyGlyLeuAlaGlu 184
542 CGGGCAGGTGGCGGATGCCCTGGTGACCTGACCGGGGCTGGCAGAAA 591
184 rGTrpAsnLeuLysGlyValAlaGlySerGlyGlnGlnAspArgPro 200
592 GATGAACCTGAAGGGGTGAGCAGAGAGCGAGGCCAGAGACAGGCCA 641
201 GLyArgTrpGlnHisArgThrCysArgGlnLeuLeuHisLeuLysAspG 217
642 GGGCGTGGGAGCAGACAGACTTGTGGCAGCTGCTCCACCTGAAGGCCA 691
217 nCysLeuIleSerCysCysValLeuSerProArgAlaGlyAlaArgGlu 234
692 GTCGCTATCATCAGCTGCTGCTGCTCAGCCCGCAGAGCGAGTCCAGAG 741
234 euGlnGluPheHisAlaPheIleValSerAspLeuArgGluLeuGlnGly 250
742 TGGGGAGATTCCATGCTTCAATTTGTTCGAGCTGGCGGAGCTCAGAGT 791
251 GluAlaGlyGlnCysIleLeuLeuLeuArgIleGlnAsnProTrpGlyAr 267
792 CAGCGGGCCAGTCACTGCTGCTGCTGCGAGATCCAGAAACCTCGGGCCG 841
267 gArgCysTrpGlnGlyLeuTrpArgGluGlyGlyGlyGlyTrpSerGln 284
842 GCGGTGCTGGCAGGGGCTCTGGAGAGAGGGGGGTGAAGGGTGAAGCCAG 891
284 aLAspAlaAlaValAlaSerGluLeuLeuSerGlnLeuGlnGlyGlu 300
892 TAGATGACGGGTGATGTGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTG 941

```

```

301 pheTrpValGluGluGluPheLeuArgGluPheAspGluLeuThrVa 317
|||||
942 TTCTGGGTGGAGGAGGAGGAGTTCTCAGGGAGTTTGACGAGCTACCGT 991
|||||
317 IGLyTrpValThrGluAlaGlyHisLeuGlnSerLeuThrGluA 334
|||||
992 TGGCTACCGCGGTACAGGAGCGCGGCACCTGCAGAGCCCTCTACAGAGA 1041
|||||
334 rGLeuLeuCysHisThrArgAlaLeuProGlyAlaTrpValLysGlyGln 350
|||||
1042 GCGTCTCTGCCATACCGCGGGCGCTGCCCTGGGGCGCTGGGTCAGGGCCAG 1091
|||||
351 SerAlaGlyGlyCysArgAsnSerGlyPheProSerAsnProLysPh 367
|||||
1092 TCAGCAGGAGGCTGCGGGAACACAGCGGCTTTCCAGCAACCCCAATT 1141
|||||
367 eTrpLeuArgValSerGluProSerGluValTyrIleAlaValLeuGlnA 384
|||||
1142 CTGGGCTGCGGGTCTCAGAACCGAGTGAGGTGTACATTGCCGTCTGCAGA 1191
|||||
384 rGSerArgLeuHisAlaAlaAspTrpAlaGlyArgAlaArgAlaLeuVal 400
|||||
1192 GATCCAGGCTGCACGCGCGGACTGGCGAGCGCGCGCGGCGACCTGGTG 1241
|||||
401 GlyAspSerHisThrSerTrpSerProAlaSerIleProGlyLysHisTy 417
|||||
1242 GGTGACAGTCATACCTCGTGGAGCCAGCGAGCATCCCGGGCAAGCACTA 1291
|||||
417 rGlnAlaValGlyLeuHisLeuTrpLysValGluLysArgValAsnL 434
|||||
1292 CCAGGCTGTGGTCTGCACCTCTGGAAGGTAGAGAAGCGGGGTCAATC 1341
|||||
434 euProArgValLeuSerMetProProValAlaGlyThrAlaCysHisAla 450
|||||
1342 TGCTAGGCTCTGTCATGCCCCCGTGGTGGCAGCGCGGCGCATGCA 1391
|||||
451 TyrAspArgGluValHisLeuArgCysGluLeuSerProGlyTyrTyrLe 467
|||||
1392 TAGCAGCGGAGGTCCACCTGGTGTGAGCTCTACCGGGGTACTACCT 1441
|||||
467 uAlaValProSerThrPheLeuLysAspAlaProGlyGluPheLeuLeuA 484
|||||
1442 GGCTCTCCACAGCACCTCTCTGAAGAGCGCGCAGGGAGTTCTCTGCTCC 1491
|||||
484 rGValPheSerThrGlyArgValSerLeuSerAlaIleArgAlaValAla 500
|||||
1492 GAGTCTTCTTACCGGGCGAGTCTCCCTAG..... 1522
|||||
501 LysAsnThrThrProGlyAlaAlaLeuProAlaGlyGluTrpGlyThrVa 517
|||||
1522 ..... 1522
|||||
517 IGLnLeuArgGlySerTrpArgValGlyGlnThrAlaGlyGlySerArgA 534
|||||
1522 ..... 1522
|||||
534 snPheAlaSerTyrProThrAsnProCysPheProPheSerValProGlu 550
|||||
1522 ..... 1522
|||||
551 GlyProGlyProArgCysValArgIleThrLeuHisGlnHisCysArgPr 567
|||||
1522 ..... 1522
|||||
567 oSerAspThrGluPheHisProIleGlyPheHisIlePheGlnValProG 584
|||||
1523 .....GTCCAG 1529
|||||
584 luGlyGlyArgSerGlnAspAlaProProLeuLeuGlnGluProLeu 600
|||||
1530 AGGTTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1579
|||||

```

```

601 LeuSerCysValProHisArgTyrAlaGlnGluValSerArgLeuCyste 617
|||||
1580 CTGACTGGTCCACATCGCTACGCCAGGAGGTGAGCCGGCTCTGCT 1629
|||||
617 uLeuProAlaGlyThrTyrLysValProSerThrTyrLeuProAspT 634
|||||
1630 CCTGCTGCAGGCACCTACAAGTTGTGCCCTCCACCTACCTGCCGACA 1679
|||||
634 hrGluGlyAlaPheThrValThrIleAlaThrArgIleAspArgProSer 650
|||||
1680 CAGAGGGGGCTTTCACAGTCACCATCGCAACACAGGATTGACAGGCATCC 1729
|||||
651 IleHisSerGlnGluMetLeuGlyGlnPheLeuGlnGluValSerValMe 667
|||||
1730 ATTACAGCCAGGAGATGCTGGGCGAGTTCCTCCAGAGAGGTCTCCGTCA 1779
|||||
667 tAlaValMetLysThr 672
|||||
1780 GGCAGTGATGAAACC 1795
|||||
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-422-869-19
seq_documentation_block:
; Sequence 19, Application US/09422869
; Patent No. 6235481
; GENERAL INFORMATION:
; APPLICANT: POLONSKY, KENNETH S.
; APPLICANT: HORIKAWA, YUKIO
; APPLICANT: ODA, NAOHISA
; APPLICANT: COX, NANCY J.
; APPLICANT: SREENAN, SEAMUS
; APPLICANT: ZHOU, YUN-PING
; APPLICANT: OTANI, KENICHI
; APPLICANT: HANIS, CRAIG L.
; APPLICANT: BELL, GRAEME I.
; TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
; FILE REFERENCE: ARCD:307
; CURRENT APPLICATION NUMBER: US/09/422,869
; CURRENT FILING DATE: 1999-10-21
; EARLIER APPLICATION NUMBER: 60/134,175
; EARLIER FILING DATE: 1999-05-13
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 2511
; TYPE: DNA
; ORGANISM: Mus musculus
; US-09-422-869-19

```

```

alignment_scores:
  Quality: 2963.00      Length: 671
  Ratio: 4.756         Gaps: 2
  Percent Similarity: 92.846  Percent Identity: 81.073

alignment_block:
US-09-768-877-2 x US-09-422-869-19 ..

Align seg 1/1 to: US-09-422-869-19 from: 1 to: 2511

1 MetArgAlaGlyArgGlyAlaThrProAlaArgGluLeuPheArgAspAl 17
|||||
109 ATGGCGGGGGTCCGGCGGAGACCGCGCGGGGAGGAGCTTCCCGGAGCG 158
|||||
17 aAlaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrProL 34
|||||
159 GGCATTCCCGGCTCGGAGCTCTCGCTCTTTTACAACCTTGTCACGCGCTC 208
|||||
34 euAlaGlnPheArgGluAspIleThrTrpArgArgProGlnGluLeuCys 50
|||||
209 TGGCCCAAGTTTCGGGAGGAGATCATCTTGGAGACGACCCCGAGAAATCTGT 258
|||||
51 AlaThrProArgLeuPheProAspProArgGluGlyGlnValLysG 67
|||||

```

1159 TCACGAGAGAGCTGCCGGAACACAGTGTCTTCCCTCGACACCCCAAGT 1200
367 eTrpLeuArgValSerGluProSerGluValTyrIleAlaValLeuGlnA 384
1209 CTGGTTACGGCTCTGGGACCCACGACGAGGTGTGGCTGTTCCTTACAA 1250
384 rGSerArgLeuHisAlaIleAspTrpAlaGlyArgAlaArgAlaLeuVal 400
1259 GACCCCGGAGCGCTTA.....GTGGCCAGACTGGGCACTGGCG 1299
401 GlyAspSerHisThrSerTrpSerProAlaSerIleProGlyLysHisTyr 417
1300 GGT.....GCCAGTCTGCACCGCGTGAACCTCCGACGAAAGACTA 1344
417 rGlnAlaValGlyLeuHisLeuTrpLysValGluLysArgValAsnL 434
1341 CCAGGCTGTGGGCTGCACATCTCGAAGGTAGCAAAACGAGAACAGCC 1390
434 euProArgValIleSerMetProProValAlaGlyThrAlaGlyHisAla 450
1391 TGCCCAAGTCTGTCTGCACCCCTGTGAGCTGCATCGCATCGCATCG 1440
451 TyrAspArgGluValHisLeuArgCysGluLeuSerProGlyTyrTyrIle 467
1441 TATGATGTGTGATATTCACATTGGGTGTGAGCTCTCACACGAGTACTACT 1490
467 uAlaValProSerThrPheLeuLysAspAlaProGlyGluPheLeuAsn 484
1491 GGGCGGCCCTAGACCTTTTGTAAAGGATGTGCCAAGGCGAGTTCCTGCA 1544
484 rGValPheSerThrArgValSerLeuSerAlaIleArgAlaValAla 500
1541 GACTCTTCTTCACTGGGAAATCTCCCTAGTGGCGGTAGAGCTGGCCAC 1590
501 LysAsnThrThrProGlyAlaIleAlaLeuProAlaGlyGluTrpGlyThrVal 517
1591 AAGGTGATCATCGCTGGAAACACCCCTGCGAGCGAGTGGAGACTGT 1640
517 LgInLeuArgGlySerTrpArgValGlyGlnThrAlaGlyLysSerArg 534
1641 GCATTTGCAGCGGTGTGTGGAGAGCTGCCACAGACTGGGGGACAGAGA 1690
534 snPheIleSerTrpProThrAsnProCysPheProPheSerValProGlu 550
1691 ACTTTCCTCTTAACCCCTGCAATCCCTGCTCCCTTCTCTGTTCCAG 1740
551 GlyProGlyProArgCysValArgIleThrLeuHisGlnHisCysArgPro 567
1741 GGTGTGGGCGCCCGGTACATCGGTATCAACCTTACAGCAAACTGGCGCT 1790
567 oSerAspThrGluPheHisProIleGlyPheHisIlePheGlnValProG 584
1791 CAGTGCACACCGACGTCAACCCATGATGGTTCCATGTCTTTCAGGTTCCAG 1840
584 LgGlyGlyArgSerGlnAspAlaProProLeuLeuLeuGlnGluProLeu 600
1841 CAGACGGTGAACAGACAGACGGGTTCCTCCGTGCTCCAGACCACTG 1890
601 LeuSerCysValProHisArgTyrAlaGlnGluValSerArgLeuCysIle 617
1891 CTAACTGTGTACCAATCGCTACGCCCGAGAAAGTGAAGCGGCTCTCTCT 1940
617 uLeuProAlaGlyThrTyrLysValValProSerThrTyrLeuProAsp 634
1941 CTTTCTGTGGGAGACTACAGATTTGTTCCCTCCACCTACTGTCCAAATA 1990
634 hGlnGlyAlaPheThrValThrIleAlaThrArgIleAspArgProSer 650
1991 CAGAGGTACCTTCAAGGTAAACATACATACCAACCAATGATAGCAATCC 2040
651 IleHisSerGlnIleMetLeuGlyGlnPheLeuGlnGlnValSerValMet 667
2041 ATCCACACCCAGGAATGCTGGCGAGGTGTCTCCAGAGAGTCTCTTAT 2090

667 talalValMetLys 671
 |||||
 2091 GGCAGTCATGAAA 2103

seq_name: /cgn_2_6/ptodata/2/lna/6B_COMB.seq:US-09-422-869-7

seq_documentation_block:

; Sequence 7, Application US/09422869
 ; Patent No. 6235481
 ; GENERAL INFORMATION:
 ; APPLICANT: POLONSKY, KENNETH S.
 ; APPLICANT: HORIKAWA, YUKIO
 ; APPLICANT: ODA, NAOHISA
 ; APPLICANT: COX, NANCY J.
 ; APPLICANT: SREENAN, SEAMUS
 ; APPLICANT: ZHOU, YUN-PING
 ; APPLICANT: OTANI, KENICHI
 ; APPLICANT: HANIS, CRAIG L.
 ; APPLICANT: BELL, GRAEME I.
 ; TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
 ; FILE REFERENCE: ARCD:307
 ; CURRENT APPLICATION NUMBER: US/09/422,869
 ; CURRENT FILING DATE: 1999-10-21
 ; EARLIER APPLICATION NUMBER: 60/134,175
 ; EARLIER FILING DATE: 1999-05-13
 ; NUMBER OF SEQ ID NOS: 30
 ; SOFTWARE: PatentIn ver. 2.0
 ; SEQ ID NO 7
 ; LENGTH: 2001
 ; TYPE: DNA
 ; ORGANISM: Human
 ; US-09-422-869-7

alignment_scores:

Quality: 2715.50 Length: 672
 Ratio: 5.252 Gaps: 1
 Percent Similarity: 76.935 Percent Identity: 76.786

alignment_block:

US-09-768-877-2 x US-09-422-869-7 ..

Align seg 1/1 to: US-09-422-869-7 from: 1 to: 2001

1 MetArgAlaGlyArgGlyAlaThrProAlaArgGluLeuPheArgAspAl 17
 |||||
 42 ATGCGGCGGGCGGGCGCGAGCGCGGAGGAGCTGTCGGGAGCGC 91
 |||||
 17 alaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrProL 34
 |||||
 92 CGCCTTCCCGCGCGGAGCTCTCGCTCTTCTCGACTTGTCTACGCCGC 141
 |||||
 34 euAlaGluPheArgGluAspIleThrTrpArgProGluGluIleCys 50
 |||||
 142 TGGCCCAAGTTCGCGAGGACATCAGTGGAGCGCGCCCGAGGATTTGT 191
 |||||
 51 AlaThrProArgLeuPheProAspAspProArgGluGlyGlnValLysG 67
 |||||
 192 GCCACACCCCGCTCTTCCAGATCACCCAGCGGAGGCGGAGGAGCA 241
 |||||
 67 nGlyLeuLeuGlyAspCysTrpPheLeuCysAlaCysAlaLeuGlnL 84
 |||||
 242 GGGGCTGTGGGGGATGCTGCTGCTCTGCTGCTGCTGCTGCTGCTGCTG 291
 |||||
 84 ysSerArgHisLeuLeuAspGlnValIleProProGlyGlnProSerTrp 100
 |||||
 292 AGAGCAGGACCTCTCGGACGAGGTCATCTCTCGGGGACAGCCGAGCTGG 341
 |||||
 101 AlaAspGlnGluTyrArgGlySerPheThrCysArgIleTrpGlnPheG 117
 |||||
 342 GCCCAGCAGGAGTACCGGGGCTCTTACCTGTCGCATTTGGCAGTTGG 391

117 YArgTrpValGluValThrThrAspArgLeuProCysLeuAlaGlyA 134
 |||||
 392 AGCGCTGGGTGGAGGTGACACAGATGACCGCTGCCGTGCGCTTGACGGA 441
 |||||
 134 rGLeuCysPheSerArgCysGlnArgGluAspValPheTrpLeuProL 150
 |||||
 442 GACTCTGTTCTCCCGCTGCCAGAGGAGGATGTGTTCTGGCTCCCTTA 491
 |||||
 151 LeuGluLysValTyrAlaLysValHisGlySerTyrGluHisLeuTrpAl 167
 |||||
 492 CTGGAAAAAGGTCTACGCCAAGGTCCATGGTCTCTACGAGCAGCTGTGGC 541
 |||||
 167 aGlyGlnValAlaAspAlaLeuValAspLeuThrGlyGlyLeuAlaGluA 184
 |||||
 542 CGGCGAGGTGGCGATGCCCTGTGTGACCTGACCGCGCGCTGGCAGAAA 591
 |||||
 184 rGTrpAsnLeuLysGlyValAlaGlySerGlyGlyGlnGlnAspArgPro 200
 |||||
 592 GATGCAACCTGAAGGCGCTAGCAGGAAGGAGGCGCAGCAGGACAGGCCA 641
 |||||
 201 GlyArgTrpGluHisArgThrCysArgGlnLeuLeuHisLeuLysAspG 217
 |||||
 642 GCGCGCTGGGAGCAGAGGACTTGTGCGCAGCTGCTCCACCTGAAGGACCA 691
 |||||
 217 nCysLeuIleSerCysCysValLeuSerProArgAlaGlyAlaArgGluL 234
 |||||
 692 GTGCTGATCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 741
 |||||
 234 euGlyGluPheHisAlaPheIleValSerAspLeuArgGluLeuGlnGly 250
 |||||
 742 TGGGGAGTTCATGCTTCATGCTCGGACCTGCGGGAGCTCCAGGGT 791
 |||||
 251 GlnAlaGlyGlnCysIleLeuLeuLeuArgIleGlnAsnProTrpGlyAr 267
 |||||
 792 CAGCGGGCCAGTGCATCTGCTGCTGCGGATCCAGAACCCCTGGGGCGC 841
 |||||
 267 gArgCysTrpGlnGlyLeuTrpArgGluGlyGlyGlyGlyGlyGlyGly 284
 |||||
 842 CGGTGCTGGCAGGGGCTCTGGAGAGGGGGGTGAAGGGTGGAGCCAGG 891
 |||||
 284 alaPalaAlaAlaSerGluLeuLeuSerGlnLeuGlnGlnGlyGlu 300
 |||||
 892 TAGATGACAGCGTAGCATCTGAGCTCTGCTGCCAGCTCCAGGAGGGGAG 941
 |||||
 301 PheTrpValGluGluGluPheLeuArgGluPheAspGluLeuThrVa 317
 |||||
 942 TTCGGGTGGAGGAGGAGGAGTTCCTCAGGGAGTTTGACGAGCTCACCGT 991
 |||||
 317 lGlyTyrProValThrGluAlaGlyHisLeuGlnSerLeuTyrThrGluA 334
 |||||
 992 TGGCTACCGGTACGAGGAGCGCGCCAGCTCCAGAGCTCTACACAGAGA 1041
 |||||
 334 rGLeuLeuCysHisThrArgAlaLeuProGlyAlaTrpValLysGlyGln 350
 |||||
 1042 GCGTGTCTGTCATACGCGGCGCTGCTGCGGGCTGGGTCAAGGGCCAG 1091
 |||||
 351 SerAlaGlyGlyCysArgAsnAsnSerGlyPheProSerAsnProLysPh 367
 |||||
 1092 TCAGCAGAGGCTCCCGGACACACAGCGGCTTCCCGACCAACCCCAATT 1141
 |||||
 367 eTrpLeuArgValSerGluProSerGluValTyrIleAlaValLeuGlnA 384
 |||||
 1142 CTGGCTGGGGTCTCAAAACCGAGTGAAGTGTATATGCGCTCTCTGCAGA 1191
 |||||
 384 rGSerArgLeuHisAlaAlaAspTrpAlaGlyArgAlaArgAlaLeuVal 400
 |||||
 1192 GATCCAGGCTGCACGCGCGGACTGGGCGAGCGCGGGCCCGGGCAGCTGGT 1241
 |||||
 401 GlyAspSerHisThrSerTrpSerProAlaSerIleProGlyLysHisTy 417
 |||||
 1242 GGTGACAGTATACCTCTGTTGAGGCCAGCAGCATCCCGGCGCAAGCATA 1291
 |||||
 417 rGlnAlaValClyLeuHisLeuTrpLysValGluLysArgValAsnL 434


```

|||||
1292 CCAGCTGTGGTGTGACCTGTGAAG..... 1319
434 euProArgValLeuSerMetProProValAlaGlyThrAlaCysHisAla 450
1319 ..... 1319
451 TyrAspArgGluValHisLeuArgCysGluLeuSerProGlyTyrTyrIle 467
1319 ..... 1319
467 uAlaValProSerThrPheLeuLysAspAlaProGlyGluPheLeuVal 484
1319 ..... 1319
484 rGValPheSerThrGlyArgValSerLeuSerAlaIleArgAlaValAla 500
1319 ..... 1319
501 LysAsnThrThrProGlyAlaAlaLeuProAlaGlyIuTrpGlyThrVal 517
1319 ..... 1319
517 IGlLeuArgGlySerThrArgValGlyGluThrAlaGlyGlySerArgVal 534
1319 ..... 1319
534 snPheAlaSerTyrProThrAsnProCysPheProPheSerValProGlu 550
1319 ..... 1319
551 GlyProGlyProArgCysValArgIleThrLeuHisGlnHisCysArgPr 567
1319 ..... 1319
567 oSerAspThrGluPheHisProIleGlyPheHisIlePheGlnValProG 584
1320 ..... GTCCGAG 1326
584 IuGlyGlyArgSerGlnAspAlaProProLeuLeuGlnGluProLeu 600
1327 AGGgTGGAGAGCCAGGACGACCCCTGCTGCTGCTGCTGCTGCTGCTG 1376
601 LeuSerCysValProHisArgTyrAlaGlnGluValSerArgLeuCysIle 617
1377 CTGAGCTGCGGCGCACATCGCTACGCCAGAGAGGAGCGGCTGTGCT 1426
617 uLeuProAlaGlyThrTyrLysValValProSerThrTyrLeuProAsp 634
1427 CCTGCTGCGAGGACCTGACAGAGTGTGCTGCTGCTGCTGCTGCTG 1476
634 hTrGluGlyAlaPheThrValThrIleAlaThrArgIleAspArgProSer 650
1477 CAGAGGGGGCTTCCACAGTACCATCGACACAGAGATTGACAGGCGCAC 1526
651 IleHisSerGlnGluMetLeuGlyGlnPheLeuGlnGluValSerValIle 667
1527 ATTACAGCGCCGAGATGCTGGGCCAGTCTCTCCAGAGAGTCTCGGTAT 1576
667 tAlaValIleLysThr 672
1577 GGCAGTGTGTAACAC 1592

```

seq_name: /cgn2_6/ptodata/2/lna/6B_COMB.seq:US-09-422-869-1

seq_documentation_block:

```

; Sequence 1, Application US/09422869
; Patent No. 6235481
; GENERAL INFORMATION:
; APPLICANT: POLONSKY, KENNETH S.
; APPLICANT: HORIKAWA, YUKIO
; APPLICANT: ODA, NAOHISA
; APPLICANT: COX, NANCY J.

```

```

; APPLICANT: SREENAN, SEAMUS
; APPLICANT: ZHOU, YUN-PING
; APPLICANT: OTANI, KENICHI
; APPLICANT: HANIS, CRAIG L.
; APPLICANT: BELL, GRAEME I.
; TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES,
; FILE REFERENCE: ARCD:307
; CURRENT APPLICATION NUMBER: US/09/422,869
; CURRENT FILING DATE: 1999-10-21
; EARLIER APPLICATION NUMBER: 60/134,175
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1
; LENGTH: 49136
; TYPE: DNA
; ORGANISM: Human
US-09-422-869-1

```

```

alignment_scores:
  Quality: 1989.00      Length: 3020
  Ratio: 3.224          Gaps: 11
  Percent Similarity: 20.430  Percent Identity: 20.298

```

alignment_block:

US-09-768-877-2 x US-09-422-869-1 ..

Align seq 1/1 to: US-09-422-869-1 from: 1 to: 49136

```

47 GlnGluIleCysAlaThrProArgLeuPheProAspAspProArgGluI 63
3810 CAGGAGATTGTGTCACACACCCGCTGTGTTCAGATGACCCAGGAGG 3859
63 yGlnValLysGlnGlyLeuLeuGlyAspCysTrpPheLeuCysAlaCysA 80
3860 GCAGGTGAAGCAGAGGGGCTGTGGGGGATGCTGCTGTGTGCTGTGCTG 3909
80 IaAlaLeuGlnLysSerArgHisLeuLeuAsp..... 90
3910 CCGGCGTCCAGAAAGACAGGACCTCTGTGACACAGTCCGGGGCCCTTC 3959
90 ..... 90
3960 CCTGTGTTGCTCTGAGCGCGGTTCTTTTGTGCTTCTCCAGCTGCTG 4009
90 ..... 90
4010 AGTACAGAGAGGCTTGGAAAGCAGAGCTGTGCCGAGCGGAGTCTCT 4059
90 ..... 90
4060 GCTGTGTGGGGAAGCAGAGAGATTCCAGAGCAGAGCTGAGGACTGC 4109
90 ..... 90
4110 ACTCTGCTCTCTGCTGAGGGGGGCTGCTTGGCTGCCAGAGAGCTC 4159
90 ..... 90
4160 CATCAGGAGGTTGGCCCTGCTGTGTGCTCTCTGACCCCGGAGTCTCAT 4209
90 ..... 90
4210 GGAGTCAGATCACACGTTTAGAATAAGAGACAATAATGCCAGCTCACA 4259
90 ..... 90
4260 GGAAGAGGGGCTGCTGCGAGAGCTTGCCTCAGATCTCTCCTCAGCTAG 4309
90 ..... 90
4310 CTCGCTGTTTCTTACAGATTGTAATATAAGTTTGCAGAAAAGTTATTA 4359

```

90 90
 4360 AACCTGTTCTGTGGGTAGACAGATACTCTGGAGGAGAAGCCCTTCTCA 4409
 90 90
 4410 GGTTCCTTACCTGGGAGTGTACCGTTTATGCTTGGCTTGTGCTA 4459
 90 90
 4460 AGTGTGCTGATTAATGACGCGGCTCAACAGTGTGACCTCATTCAGAT 4509
 90 90
 4510 TTCACATGTCCTCCAGGCCCATGTAAGCGTGCACAGTCACTGCGCTTT 4559
 90 90
 4560 CAGACACATGGTCTTACCAGCTTTGACTTTTTTTTAAACGAGAGTCTA 4609
 90 90
 4610 AATCACTGCCATTGTGTTTCTGGCGTAAAGTGGCAGAGCCAGACCGC 4659
 90 90
 4660 ACCAGGTGCCGTGCCAGCCTGCACTCCCGCATGCTGGGTCAAGATCC 4709
 90 90
 4710 TTACCCCTGAAGAGCCCTGCGGTGGAGCGCTGTGGTGCAGCAGTGGC 4759
 90 90
 4760 CCAGAGTCGGGGCCGAGCTCCAGCAGCAGGAGGGCTGTGTTCTCG 4809
 90 90
 4810 TGGTGACGTGTGCTTGACGAGCTCGGTCAAGAACTGGGTCACTCATG 4859
 90 90
 4860 CCTTGAATGTCACATTTGTTTGGCTTCAGTCAAGTCTTTTAGTGAG 4909
 90 90
 4910 GGCAGCAGAGTGTCTCCGGGATATGTGGCTCCCTCGGTGTGGTCTCAA 4959
 90 90
 4960 GTTTTCAATGAGAGTGTGTTAATTCATGTGGGTGATGCAGCCCTGTG 5009
 90 90
 5010 CAGGCGCCGACATCCAGGTGTCCGTAGAGTTCTCTGGACATCCAGGTG 5059
 90 90
 5060 TGGGTGAGAGTTCTCTGGACATCCAGGTGTCCGTAGAGTTCTCTGCGA 5109
 90 90
 5110 CATCCAGGTGTCCGTAGAGTTCTCTGGACATCCAGGTGTGCGTTAGAG 5159
 90 90
 5160 TTCTCTGCAGACGGGTGCTGTGGAGCACTCAGCTGTGCCACACCGC 5209
 90 90
 5210 GCGCGGACACTGGAGTAGCGCGGTGTGCTTATATCACGCTCGCCTT 5259

91GlnValIleProGlyGlnProSerTrp 100
 5260 TTGCTTCCTCCCTGTGCATGGCAGGTCAATCCCTCCGGGACAGCGAGCTGG 5309
 101 AlaSpGlnGluTyxArgGlySerPheThrCysArgIleTrpGlnPheG 117
 5310 GCCGACAGGAGTACCGGGCTCCTTCACCTGTGCGCATTTGGCAGTTGG 5359
 117 yArgTrpValGluValThrAspAspArgLeuProCysLeuAlaGlyA 134
 5360 AGCTGGGTGGAGGTGACACAGATGACCGCTGCGCTGCCTTGCAGGA 5409
 134 rLeuCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
 5410 GACTCTGTTCTCCCGCTGCCAGAGGGAGATGTGTTCTGGCTCCCTTA 5459
 151 LeuGluLysValTyralaLys 157
 5460 CTGGAAGAGTCTACGCCAAGTGCCTGTGCTGGGGCTGAAGGGCTGCGC 5509
 157 157
 5510 CTGGGCAAGTGGGAGCTGCCACTACCATGGGTGCCCCAGGAGGCTCTC 5559
 157 157
 5560 TGCTCACTCTGGCTGCAGAGCCCTTCAGTTCCTGAGGGTCTGGCAGCT 5609
 157 157
 5610 CATCTGTGAGTCAGGCTGACAGGCCAGGTGCAGAGATTCTTCTTTTGG 5659
 157 157
 5660 CCTGTGGATTGCCACTCCCTGCTTCCCTTCCCTTCCAAAGCCAG 5709
 157 157
 5710 CGTGGAGTGTTCCTCCAGAGAACATGTGTGCGCTCCTCTTATTTAT 5759
 157 157
 5760 CGGCCCCAGCAAGAAGATGCTTCTTATTTGTTGTGGAATGCTTGGG 5809
 157 157
 5810 ACAGCAGACTCATTTGTGTAGTGTGGGAGGAGTGAGGTACCCAGC 5859
 157 157
 5860 ATACCACACTTGTGTATCACGGTGTGCTGGCTCAGGGGACAGGACC 5909
 157 157
 5910 CTCACCATGAGTCATAATTGAATAGCCTTCCCTCTTAGAATGCATTTGTC 5959
 157 157
 5960 TTCCTGCCAAAGGCACTGGAGTGCAGGAGGAGGAGGAGCTGTGTAAC 6009
 157 157
 6010 ATGGAGGCTGGCTGGTGACATCAGTGCCTCCAGTGCAGCCCTTCCATCCCA 6059
 157 157
 6060 AGGCTGTTTTAGGAAAAGCAGGTTGGAGCTTGAGAGCAAGGATGTG 6109
 157 157
 6110 GGCATCCATAGCTCCACGCTCCTGCGCTGTGCTGTGTCCTGTGTCACACCGGA 6159
 157 157

6160 TGCCAGAGAGTTTCTGTGTGTGGCAGAGACTGCAGGGCGCTCAGCCTT 6209
157 157
6210 GCTGCGAAGTAAGCGCTTTGAAGGTGAGGCTAAGCCTTGACTGTGTGAG 6259
157 157
6260 ATGAGGAAGAAAGCAGAGGGAGTAAGAGGTGGGATTGAGCAGCGGT 6309
157 157
6310 GGAGATTGGGGTGTCTACAGACATGGGAATCAGAGAGGGGCCATGCT 6359
158 ValHisG1 160
6360 CAATGCCAGAGGCTCACTCCCATGTGATGTGTCCCTAGGGTCCATGG 6409
160 ySerTyrGluHisLeuTrpAlaGlyGlnValAlaAspAlaLeuValAspL 177
|||||
6410 GTCTTACAGACCTGTGGGCGGCGAGGTGCGGATGCCCTGTGTGGACC 6459
177 eutHrGlyGlyLeuAlaGluTrpAsnLeuLysGlyValAlaGlySer 193
|||||
6460 TGACCGGCGGCTGTGCGAAGAGATGAACTGAAAGGCGGTAGCAGGAAC 6509
194 GlyGlyGlnGlnAspArgProGlyArgTrpLysHisArgTrpCysArgL 210
|||||
6510 GGAGGCCAGAGGAGCAGAGGCCGCGCTGGAGCAGACAGACTTGTGGCA 6559
210 nLeuLeuHisLeuLysAspGlnCysLeuLysCysValLeuSerP 227
|||||
6560 GCTGCTCCACTGTAAGGACAGTGTGTGATCAGCTGCTGCTGCTCAGCC 6609
227 roATrGAlaGly 230
6610 CCAGAGAGGTGAGGACGCTGGCCACATGGAGGGCTGCAGCCAGCGTG 6659
230 230
6660 CCCCCACATGCGCAGGCTCAAGGCACACTGATTTTATGTGACTGGCT 6709
230 230
6710 ACACAGCCCTGTCAGACTAAGTGGGAAGTAAGCTGTCTCAAGGG 6759
230 230
6760 TGGTGTCTCAGTTTGTGACCTTCCCTGCTGTCTCTCCAGAGGAGC 6809
230 230
6810 TGSCCCTTCTCCCTGACCAGTCTTTCACACTAGTGCAGAGCAGAG 6859
230 230
6860 AGGTGACCGAGCTCAAGCCCACTGTCTGTGCCATCCCTGGCCAGCTG 6909
230 230
6910 GCACCTGGCAAAATCAAAACCTGTTTTTATTTAGTATAGATACAT 6959
230 230
6960 TCGTAAAAACAGTTTGTCTCCAAAAATGAAGGGCCAGGTGTGTGG 7009
230 230
7010 CTCACACTGTATCCAGAACTTTGGAGGCTGAGGTGGAGTATCGCT 7059
230 230

7060 GGAGCCAGAGATTCAAGAACCCGCTGGGCAACATGGCAAGATCTCAT 7109
230 230
7110 CTCTACAAAAAATGAAGAAAAAATCATAGATGAACCATGTGAC 7159
230 230
7160 TTTTACGTCCGCTCTCAGTTTCCATGACACAGCAGGTTTACGTGTCC 7209
230 230
7210 CTCCGACCGCTTCTCACACTGCCACAGAGCGCTGTCAATGTGCC 7259
230 230
7260 CCATGAGCGGTTTGCACAGCTCTTATATCTTCCCTGATGTGTGATA 7309
230 230
7310 TTAGATCTCTGTGTTTATATATATTAAGCAGTATCAATGAACATCT 7359
230 230
7360 TTATCATTTTTCATACTTAGATTAATTTAAATATGGTTAAAGAT 7409
230 230
7410 ATGATATCAGAGTAACTAACAAAGTTGTATAGGCTGTGTGGGT 7459
230 230
7460 CCCCCATAGACACCTGAGATAGTTTCAGAGCAGCTTGAGAGGTTTC 7509
230 230
7510 AGAGAGCTGGAGGGGCTGTACTCCCGCCAGCCGCTGGGGTGCCAGG 7559
230 230
7560 AAGAGCCTGGCTCAGAGCCTCCACCTGGAGTGAAGGGCTGTGTG 7609
230 230
7610 GTGGCATTAAGTGGGAGCTGTGGTTTGGTGCTCCAGGGTGGGT 7659
230 230
7660 AGCGTAAGTCTAGCACTTCAGGCTTTAGAGGAAGCCCCCAGCAGAGA 7709
230 230
7710 GAGATGCACCTGGCACTGACTGAGGTGCATGAGACCTGTAGAGTGG 7759
230 230
7760 CAAGGGGCGGCGTGGGCGAGACTGAAGTCTTCCAGAGACCTCTCT 7809
230 230
7810 CTCTAGACCTTGCCCGCAGAGATGTGCTTCCCTCTCTGTGTGGCTC 7859
230 230
7860 TTTCTTTCTATGCCCATCTGTCTCAAGTCTGTGGCAGAGTGAAGCTGT 7909
230 230
7910 TGCCCCCTGGGCTCCAGCTCCAGTGGGCTTACAGCAGGCACTAT 7959
230 230
7960 CTGATGTCTACTCTAGAGGCTTTGCCAAGAAACACTCTAGAT 8009

230 230
8010 TTCTGCACTGTCAGATGATCTGTGCAGCATGTTTGAGAACTGGTGCTG 8059
230 230
8060 TGCCAGGCATGCTGTCACAGATGGGTACGGACGATGACTGAGCCCTAGAGA 8109
230 230
8110 AGGGGATGACTTACCAGCACCAGACCCGGATGGCAGCCGATCAGGCC 8159
230 230
8160 CGTGTGCTTCTCTGGAGATGCTCTGAGCTGATGGGTACAGCTGCTA 8209
230 230
8210 AGAAGAGACTCTGTGTGTCCATGAGGATCTGGGGTCTCCCTGTGTCG 8259
230 230
8260 TAGTGGGTGGCTGGAGCATCTCCAGAGAGGAGGACGAGCCCTGTGTG 8309
230 230
8310 TCCTTGTCACTTTGGGACTCCATGTTGTCCTTCTGCTGCTCCGCC 8359
231 Ala: ArgGluLeuGlyGluPheHisAlaPheIleVa 242
8360 ATTCTCATGCAAGTGCCTGGGAGCTGGGGAGTTCATGCTTCTATGT 8409
242 lSerAspLeuArgGluLeuGlnGlyGlnAlaGlyGlnCysIleLeuLeuL 259
8410 CTCGGACCTGCGGAGCTCCAGGTCAGGGTCAGGGCCAGTGCATCTCTG 8459
259 euArgIleGlnAsnProTrpGlyArgArgCysTrpGlnGlyLeuTrpArg 275
8460 TCGGATCCAGAACCCCTGGGGCCGGGTGCTGGCAGGGCTCTGGAGA 8509
276 GluGly..... 277
8510 GAGGGGTGAGTCTGGGGCTGGACCATTGCTGCTCGGGAGGGGGCCC 8559
277 277
8560 AGTGCCAGTCTGGCCTGTGCTCTGCTGCTCACCTTCAGCTGTGAGGACTGTAC 8609
277 277
8610 TTGGCTGTCTCAGCAAGGCCCTCAGTCCCTGCTCTGCTGACACCATGC 8659
277 277
8660 TTGCTTGGCTCCAGGCAATCTTGTGAGGCTGGGACCAGAGTGGCCAT 8709
277 277
8710 TGGGCTTGGGTTTCAATAGGCGACACATCATCAGGGCTCGGGCAGCAG 8759
277 277
8760 TCCTGGGAAGACGATCCAGAGCGCTGAAGTTCTCTGGGAAAAGAGAGG 8809
277 277
8810 GCTCCAGGGTGGCCCTGCCAGAGAGGCCCTGTGCTGCAGAGCTGCTTCG 8859
277 277
8860 GGTGTGGAGGGCTGCAGAGCTGGGGCAGCGGGTCTGCTGCAGGAACTC 8909

277 277
8910 AGGGCTCTCTGGTCCCTCCAGGCTTCCCCCAGCCTGCCGCAAGTTG 8959
277 277
8960 ACACCTACCAGTTCTCGGAGGGCTTCTCTGAGATGAGGTTTCTTCCAG 9009
278 .GlyGluGlyTrpSerGlnValAspAlaAlaValAlaSerGluLeuLeuS 294
9010 GGGTGAAGGTTGAGCCAGGTAGATCAGCGGTAGCATCTGAGCTCTCTGT 9059
294 exGlnLeuGlnGlyGluPheTrpValGluGluGluGluPheLeuArg 310
9060 CCCAGCTCCAGGAAGGGAGTTCTGGGTGGAGGAGGAGTTCCTCAGG 9109
311 GluPheAspGluLeuThrValGlyTrpProValThrGluAlaGlyHisLe 327
9110 GAGTTTGACGAGCTCACCGTTGGCTACCGGTCACGAGGCCGC .CACCT 9158
327 uGlnSerLeuTyThr..... 332
9159 GCAGAGCCTCTACACAGGTAGTGCCTCCGAGGGGCTGTGCTGGGACGTGC 9208
332 332
9209 TCTGCTGCGGAAGTGAAGGAGGCTGGGCACGTTGCTGGTTCCTCCCTGC 9258
332 332
9259 CCAGGCCAGTTTGGTTCTTTCAGCGTGGAGAGATGATTCTGTCTCCAGG 9308
332 332
9309 AGCCGGGAGGAGGTGATGATTCTGTCCAGGAGCTGGGAGGAGGTGGG 9358
332 332
9359 CTGTGTGGAGGGCTGGCTCTGCTGTGCGCTAGTCTGTCTTAGACCC 9408
332 332
9409 TCACCAGGTTTCATGAGGCCACCGTGGCGGAGGCCAGCAGGAGCGGTGT 9458
333 GluArgLeuLeuCysHisT 339
9459 CCCACAGCTGATGCTCTGGTGTCTTCTACTAGAGAGCTGCTCTGCCATA 9508
339 hrArgAlaLeuProGlyAlaTrpValLysGlyGlnSerAlaGlyCys 355
9509 CGCGGGCTGCTTGGGGCTGGGTCAAGGGCCAGTCAGCATGAGGCTGC 9558
356 ArgAsnAsnSerGlyPheProSerAsnProLysPheTrpLeuArgValSe 372
9559 CGGAACAACAGCGGCTTTCACAGCAACCCCAAAATTCCTGGCTCGGGTCTC 9608
372 rGluProSerGluValTyrIleAlaValLeuGlnArgSerArgLeuHisA 389
9609 AGAACCGAGTGAGGTGTACATTCCTGCTGCAGAGATCCAGGCTGCAGG 9658
389 laAlaAspTrpAlaGlyArgAlaArgAlaLeuValGlyAspSerHisThr 405
9659 CGCGGACTGGGCAAGCCCGGGCCAGTGGTGGGTGACAGTCTACT 9708
406 SerTrpSerProAlaSerIleProGlyLysHisTyrGlnAlaValGlyLe 422
9709 TCGTGGAGCCCGAGGAGCATCCGGGCAAGCACTACCAGGCTGTGGTCT 9758
422 uHisLeuTrpLysVal..Glu..... 428
9759 GCACCTCTGGAGGTTACTACCCCTGCTGGCTCACGCTCGGTTCAGCA 9808
428

9809 GGTGTGTGAGGCCCATGAGCTCTGGGTTCTAGACTGGCTCTGCCG 9858
 428
 9859 GACACATGTACTCTGCCACGGGCCCAACAGTCTCCCTCTTTGGC 9908
 428
 9909 TGTGCAAGGGGTGAGCTCTGCTGCTCCACAGCCCGCTGTACT 9958
 428
 9959 GAGACTCAGGCTTCTGCTCTGCTGCTCCATTTGGCCAAAGAA 10008
 428
 10009 GCAACAGATCTGACGACGAGGAGACCCGACAGAGGGCTGACCCG 10058
 428
 10059 GGGGGCTCTGGACGCCCTTCTCTAGGCAAGAGGTAGAGCCAG 10108
 428
 10109 CGGCTGAGACCTGTCAAGGCCAGTCCAGCTCTGACGTTGCTGTGA 10158
 428
 10159 CTGGACACATCTCTCCCTGCTCCCTCAGTCTTCCCTGACAGAC 10208
 428
 10209 GGGGTCTGACAGGATCTATGGGATGCTGTAGGCCACAGCATCCC 10258
 428
 10259 AGGCTCAACCATGCTTCAACAAAGTGTGTTCCAGAAACAGATG 10308
 428
 10309 GGGGGCTGAGGGCAAGGGCTGAGTGTGAGAGATATGCCGCTG 10358
 428
 10359 CTGCTCAGGGGCTGGGTTTTCATCTTGTGTCTTGACAGGGGTGACA 10408
 428
 10409 CTTGGACACACTGTTCCCTGCTCATGATGTGGCCACATGATG 10458
 428
 10459 TTCCTTCTCTTGCAAAAGAGTGTGAGAGGCCACTGTCCACAGC 10508
 428
 10509 CCCAGGTTGCTGGGCCACAGTGCTTGTGGGCCACGTTACAGAGAG 10558
 428
 10559 ACTTGACAGCTGTCTGTGGACAGATACTGGGCCAAGGCAATGAAG 10608
 428
 10609 CCCGGATTGTGGCATCTTAGCTGTCTGAGAGAGGTGAGGGT 10658
 428
 10659 GGTGACAGGCTTGGGCTTTTCATCTGTCACTCCAGAGGCCCTTGTCT 10708
 428

10709 TGCACAGAGAGTCAAGGCCGCTGTGACCCCGCCCTCTCTCAC 10758
 429
 10759 ACTGAGCTCTGACAGTGTCAAGGTAGAGAAAGCGGGTCAATCTG 10808
 435 ProArgValLeuSerMetProProValAlaGlyThrAlaCysHisAlaTy 451
 10809 CTTAGGGTCTCTCATGCTCCCGCTGGCTGGCACCGTGTCCATGCA 10858
 451 rAspArgGluValHisLeuArgCysGluLeuSerProGlyTyrTyrLeu 468
 10859 CGACCGGAGGTCACCTGCTGTGAGCTCTCACCGGGCTACTACTGG 10908
 468 LaValProSerThrPheLeuLysAspAlaProGlyGluPheLeuArg 484
 10909 CTGTCCACAGACCTCTCTGAAGACGGCCAGGGAGTTCTCTCTCGA 10958
 485 ValPheSerThrGlyArgValSerLeu..... 493
 10959 GTCTTCTACCGGGCGAGTCTCTCTAGTGTAGAGAAACCGCAGTGC 11008
 493
 11009 TGTGCTCTCCAGGCCACAGGCCCTTCCAGCAGAGATTGGCACTT 11058
 493
 11059 TCCCTGTGTGTTGGCAGGTGTCCATGTGGAACTGAGGCCACTGGAC 11108
 494
 11109 CTGCTGCCAGCCCTCCATGTTTGTCTTCTTGGCAGCGCATCAGGCG 11158
 498 aValAlaLysAsnThrThrProGlyAlaAlaLeuProAlaGlyLuprg 515
 11159 AGTGGCCAAAGACACACCCCGGGGCAAGCTCTCTCGGGGAGTGG 11208
 515 LThrValGlnLeuArgGlySerTrpArgValGlyGlnThrAlaGly 531
 11209 GACCGTGCAGCTACGGGTTCTTGAGAGTGGCCAGACGGGGGGGC 11258
 532 SerArgAsnPheAlaSerTyrProThrAsnProCysPheProPheSerVa 548
 11259 AGCAGGAATTTTCTCTATACCCCAACCCCTGCTTCCCTCTCTGGT 11308
 548 LProGlyLysProGlyProArgCysValArgIleThrLeuHisGlnHis 565
 11309 CCCGAGAGGCGCTGGCCCGCTGCTGCTGCTGCTGCTGCTGCTGCT 11358
 565 ysaArgProSerAspThrGluPheHisProIleGlyPheHisIlePhe 580
 11359 GCCGGCCACTGACACCGAGTTCACCCCATCGGCTTCATCTTCCAG 11408
 580
 11409 GCAAGCTCTTGCCCCAGGAGGAGGAGGAGAGAGAGGGCCCTCAGAG 11458
 580
 11459 AATTGATCTTGCGCTCCATTTGTCACAGAGAGGCTCTGGCTCAGTC 11508
 580
 11509 ACTTGGGCTCCCTGCTTGGAGGCGTGCCTAGAACCCGACAGGGC 11558
 580
 11559 CTTCTCCATCTCCAACTCTCAGAGGCAAGGCCAAGATGGCTGTGA 11608
 580
 11609 AGGGCGGGGCGCTGGAGGTGGGCAAGGCTGATCCAGGAGGCAAGTT 11658

580 580
11659 TCCAGAGAGGTGCTGAGTGGGAGGAGGAGAGTCTTGGAGAGCAG 11708
580 580
11709 GAGGCGAGGTTGAGACAGCGGGGTGGTTCAGCCCTGGCTTGGGAAC 11758
580 580
11759 GCAGGGGCTGATGGACTCAGGAGTGCAGAGGGGAGGCCAGGCTGGC 11808
580 580
11809 TGSCCAGACAGCCCTCGGTGTGAGGAAGTCCACAGTCACTGAGCTCA 11858
580 580
11859 GCCAGAGCCCTGTCCACTTACCCTGACTCAGAACTGCTGCTCCCAAG 11908
580 580
11909 GTTCATTCTGCAGACATGTGTCCCTGGATGCAGGGGCCCTGACGAG 11958
580 580
11959 GAAGGCACTGCAACCTCGGTTCACAGTGGGTGCTGGGACCTTTGGA 12008
580 580
12009 CCTCGCTGTTTGGCCTGGGCCACCGGCTCAGGTCCCTAGAGCTGTAG 12058
580 580
12059 GAAACACATGCCAGGGCCAGTGGGAGCCCTTGGGGCGGGCTGGGCGAGTC 12108
580 580
12109 ACAGGTGTTAAGCCCTCATGATGTGACAGGCTCCAGCGGGGGCCCC 12158
580 580
12159 ACTGCGGCACCTTTGCAAGGGTGGCCAGGCCTTGTGAGGAGCGAG 12208
580 580
12209 TCCAGTGTCCAGCCCTGGCGAGCCCTCTCAGAGAGGGGCTGTATGCA 12258
580 580
12259 CTCAGAGGGCCAAAGGCATCCGAGCAGATGCCCTGGGCTGGGCTCCCT 12308
581 GlnValPr 583
12309 ACCCAAGGCTGGCCCCCTCAGTCTGAGCCCTGCGCTTTCCCTCAGGTCCC 12358
583 oGluGlyGlyArgSerGlnAspAlaProLeuLeuGlnGluProL 600
12359 ACAGGTTGAGAGGAGCCAGGAGCCACCCCTGCTGCTGAGGAGCCGC 12408
600 euleuSerCysValProHisArgTyrAlaGlnGluValSerArgLeuCys 616
12409 TGTGAGCTGCGTGGCCACATCGTACGCCAGGAGGTGAGCCGCTCTGC 12458
617 LeuLeuProAlaGlyThrTyrIysValValProSerThrTyrLeuProHis 633
12459 CTCTGCTGCGGGGACCTTACAAAGTTGTGCCCTCCACCTACTGCCGGA 12508
633 pThrGluGlyAlaPheThrValThrIleAlaThrArgIleAsp..... 647
12509 CACAGAGGGGGCTTCACAGTGCACCATGCCAACAGGATTGACAGTGGG 12558

647 647
12559 GCTCTGGGACTTGGGGCGGCCAGCTGGAGGCTGGGGTCTGTGAGTCTTA 12608
647 647
12609 GTGCTGCGCTGTCCCGCCACAGTCTCTGCTGCTGCCCTCACCCCTCAAGCCC 12658
647 647
12659 CTATCTGTCTGCGCAGACAGGCGCTCTCTGCTACCTGGGGACCCCTTC 12708
647 647
12709 TTCTGCTGTGAGCCCTGGAAGAGAGTCTAGTGGGAGGTGGGCCAGGAGC 12758
647 647
12759 ACACAGCCACTTGTGTGCAAGTGCAGTCTGGGAGCGCTGATCTGGTGTG 12808
648 ArgProSerIleHisSerGlnGluMetLeuGlyGlnPheLeuG 662
12809 TCTCCACAGGCCATCATTACAGCCAGGAGATGCTGGGCCAGTTCCTCC 12858
662 InGluVal 664
12859 AAGAGGTG 12866
seq_name: /cgn2_6/ptodata/2/1na/5B_COMB.seq:US-08-835-099A-5
seq_documentation_block:
; Sequence 5, Application US/08835099A
; Patent No. 5874277
; GENERAL INFORMATION:
; APPLICANT: SHINTANI, Yasushi
; APPLICANT: NISHI, Kazuori
; APPLICANT: KAWAMOTO, Tomohiro
; TITLE OF INVENTION: NOVEL PROTEINS, THEIR PRODUCTION
; TITLE OF INVENTION: AND USE
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/835,099A
; FILING DATE: 04-APR-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 083649/1996
; FILING DATE: 05-APR-1996
; APPLICATION NUMBER: 97105508.2
; FILING DATE: 03-APR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Resnick, David S
; REGISTRATION NUMBER: 34,235
; REFERENCE/DOCKET NUMBER: 47342
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6440
; TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2109 base pairs
; TYPE: nucleic acid

STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 US-08-835-099A-5

Alignment_scores:
 Quality: 635.00 Length: 508
 Ratio: 2.096 Gaps: 17
 Percent Similarity: 59.646 Percent Identity: 32.283

alignment_block:
 US-09-768-877-2 x US-08-835-099A-5

Align seg 1/1 to: US-08-835-099A-5 from: 1 to: 2109

```

13 Leupheargaspalaalaphroalaalaspserleu...phecy 28
|||||
133 CTATTAGAGACCTGAGTCCACATGTCATGAGCTTGAGCTACAA 182
28 saspleuSerthrProleualaglnphearglunspilthrParga 45
|||||
183 GGAATCTTGA...CCAGCTCTCCCAACATCAAGCATCATCTGGAAGC 229
45 rProglnglnleucysalathrProargleupheProaspasProarg 61
|||||
230 GGGCCAGGAGTGTGTCCAGCCCTCAGTTATCTGTGTGAGCCACG 279
62 gluglgnVallysglnglnleuenglyaspCystrpheleucysal 78
|||
280 CGCACAGACATTTGTCAGGTGTCTAGAGTACCTGCTTGCTTGCTGC 329
78 acyalaalaleuclnlyseratglnleuaspalnvalilleprop 95
|||
330 CATTCCTCCCTGACCTGAAATGAAGCTCTTACCGGGTGGCCCA 379
95 roglnglnProserThrpalaspnglnlytrarglyserPheThrCys 111
|||
380 GGGACACAGAC...TTCCAGAGACATATGGGGAAATCTTCACTTT 423
112 Argllletrpglnphegllyargtrvalgluvalthrthrasparglye 128
|||||
424 CAGTTCTGGCAGTACGAGAGTGGTGGAGTGTGATTCAGCAGAGCT 473
128 uproCysleualaglyargleucyspheSeratgCysgluArglunspv 145
|||||
474 GCGCACCAAGAAATGACAGCTGCTTCTTACCTCGAACAAGCAATG 523
145 alpheTrleuproleuclnlyserValtrAlalysvalHISglySer 161
|||||
524 AATTCTGAGTGCCTCTGAGAAAGCTTATGCAAGCTTAAAGGTGTG 573
162 TyrglnHISleuThrpalaglygluValalaspalaLeuValaspleuth 178
|||||
574 TATGAGGCTCTGCGTGGAGTTCACAGTGGAGGCTTTGAGGATTTTCA 623
178 rgluglyleualaglyargtrpansleuylgluValalaglysergly 195
|||||
624 AGGTGGATCTGTGATTTTATGACCTGAG... 654
195 lylnglnlnasparpProgllyargtrpelnHISargthrcysarglnleu 211
|||||
655 .....AAACCAACAGCAATCTATATCATGATCATCCGAGAGGCC 693
212 leuHISleuLysarglnCys.....leuHISerCysCysVal.. 224
|||
694 CTC.....TGTGGGGGTCTCTGCTGGGCTCTCCATTTGA 728
225 leuSerProalaglyala.....ArggluLeuLys 236
|||||
729 TGTCTCCAGTCCAGCCGAGCCGAGACCATCACACAGCAAGACTGGTTA 778
236 lvrheHISalaphelleValserAspleuarggluLeuenglnglnala 252

```

```

779 AGAGTCATGCTACTGTCTGACTGGAGTGCAGAAAGTGAATTTCCAG... 825
253 GlylnCysIleleuLeuValargllleglnasnProtrpOlyargargcy 269
|||||
826 GGCATCCAGAGAGCTGATCAGACTCAGAAATCCATGAGGGTGAAGTGA 875
269 strpglnglyleuThrpargluglygluLyslytrpsergluValaspa 286
|||||
876 GTGTGCGGAGCTTGAGCGATGATCACCAGAGTGAATCAATGAGACC 925
286 lAlaValalasegluLeuSerGlnleuenglnglnlyglupheTrp 302
|||||
926 CCCGCGGAGAGAGAGACTGACAGCAAGAAATGAGATGAGAAATTTCTG 975
303 ValgluglngluGlupheleuargglupheaspgluLeuThrValglly 319
|||||
976 ATGTACTTTCAGATTTCTGAGCAGATCTCTGCTGGTGAAGTCTGCA 1025
319 rProvalThrualaglyHISleuGlnserleuLythrGluargleu 336
|||||
1026 CCGTCCCGGAGCTCTGAGTAGCGAGAGGTGCACAAATGAACTTGG 1075
336 euCysHISThrargAlaleuProgllyAlaTrpVallysglnserAla 352
|||||
1076 TCCTGTTCAC.....GGCACAGTGCACCCGGGGCTCCACAGCT 1113
353 GlylyCysArgasnSerGlyphePro.....SerAsnPr 365
|||||
1114 GGGGGCTCCAGAAC.....TACCAGCAGACTGATGACCAATCC 1154
365 OlyspheThrleuargValserGluProserGlu..... 376
|||||
1155 CCAGTCAAAATCCCTTGGATGAAGTGAAGAGACAGAGAGAGACA 1204
377 .....ValTrillealavalLeuGlnargserArg 386
1205 TCGGTGAACCTGCTGTACAGTGTCTGGGCTGTATGCAGAAATATGC 1254
387 leuHISalalaspThrpalaglyargAlaargAlaleuValglyaspse 403
|||||
1255 .....AGGTGGCGGAAGCCGAGATGAGCAAGG 1280
403 rHISThrSerThrProalaserIleProglyLysHISyrglnAlay 420
|||||
1281 CATGCTTAC.....A 1291
420 alglyleuHISleuThrplysValgluysarg.....Arg 431
|||||
1292 TCGGCTATGCCGCTACAGGTTCCAAAGAGCTGGAGAGTACACAGAC 1341
432 ValasnleuProarg.....ValleuSerMetProvalAlaglyTh 446
|||||
1342 GCACACTTGGCCCGGATTTCTTCTGCTTACAGCCCTCAGCCGAGC 1391
446 rAlaCysHISAlaTyAspargluValHISleuArgCysgluLeuSer 463
|||||
1392 CAGACCTTACGTCAACTCGGGAGGTCTGTGGCGGGCCCGGCTGCC 1441
463 rogllyTyTrleuAlaValProserThrPheleuLysAspAlaProgly 479
|||||
1442 CTGGGAGTACCTGTGTGCTGCATCATTTGAACCTTCAAGAGAGCC 1491
480 GlupheleuLysarglyalPheSer 487
|||||
1492 GAGTTCTGTGAGAGTGTCTCA 1515
seq_name: /cgn2_6/ptodata/2/lna/6A_COMB.seq:US-09-157-349-5
seq_documentation_block:
; Sequence 5, Application US/09157349
; Patent No. 6068990
; GENERAL INFORMATION:

```

APPLICANT: SHINTANI, Yasushi
APPLICANT: NISHI, Kazuori
APPLICANT: KAWAMOTO, Tomohiro
TITLE OF INVENTION: NOVEL PROTEINS, THEIR PRODUCTION
TITLE OF INVENTION: AND USE
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/157,349
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/835,099
FILING DATE:
APPLICATION NUMBER: 97105508.2
FILING DATE: 03-APR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Resnick, David S
REGISTRATION NUMBER: 34,235
REFERENCE/DOCKET NUMBER: 47342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
TELEX:
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2109 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-157-349-5
alignment_scores:
Quality: 635.00 Length: 508
Ratio: 2.096 Caps: 17
Percent Similarity: 59.646 Percent Identity: 32.283
alignment_block:
US-09-768-877-2 x US-09-157-349-5 ..
Align seg 1/1 to: US-09-157-349-5 from: 1 to: 2109
13 LeuPheArgAspAlaPheProAlaAlaAspSerLeu...PheCy 28
|||||:|||||
133 CTATTAAAGGACCCCTGAGTCCAGCATGCCATCAGCTTTGGGCTACAA 182
28 sAspLeuSerThrProLeuAlaGlnPheArgGluAspIleThrTrpArgA 45
|||||:|||||
193 GGATCTTGG...CCAGGCTCTCCCAAACTCAAGGCATCATCTGGAAGC 229
45 rGProGlnGluIleCysAlaThrProArgLeuPheProAspProArg 61
|||||:|||||
230 GGCCACGAGGAGTTGTGCCAGCCCTCAGTTATCGTTGGTGGAGCCACG 279
62 GlnGlyGlnValLysGlnGlyLeuLeuGlyAspCysTrpPheLeuCysAl 78
::: ::::: |||||
280 CCACAGACATTTGTCAAGGGTGGTCTAGGTGACTGCTGGCTTCTGGCTGC 329
78 acysAlaAlaLeuGlnLysSerArgHisLeuLeuAspGlnValIleProp 95
| |||||:|||||

330 CATTGCCTCCCTGACCCCTGAATGAAGAGCTGCTTTACCGGGTGGTCCCCA 379
95 roGlyGlnProSerTrpAlaAspGlnGluTrpArgGlySerPheThrCys 111
||| :|||:|||||
380 GGGACCCAGGAC.....TTCCAGGAGAACTATGCGGGAATCTTTCATTT 423
112 ArgIleTrpGlnPheGlyArgTrpValGluValThrThrAspAspArgLe 128
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
424 CAGTTCTGGCAGTACGGAGAGTGGGTGGAGGTGGTCAATTGACGACAGGCT 473
128 uProCysLeuAlaGlyArgLeuGluCysPheSerArgCysGlnArgGluasp 145
||| :|||:|||||
474 GCCCACCAGAATGGACAGCTGCTCTCTCTACACTCGGAACAAGCAATG 523
145 alpheTrpLeuProLeuLeuGluLysValTrpAlaLysValHisGlySer 161
||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
524 AATTCTGGAGTCCCTGCTGGAGAAAGCCATATGCCAAGCTTAATGGTTGT 573
162 TyrGluHisLeuTrpAlaGlyGlnValAlaAspAlaLeuValAspLeuTh 178
||||| ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
574 TATGAGGCTCTCGTGGAGTTCACAGTGGAGGGGTGGAGATTTTCAC 623
178 rGlyGlyLeuAlaGluArgTrpAsnLeuLysGlyValAlaGlySerGly 195
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
624 AGGTGGCATCTCTGAGTTTATGACCTGAAG..... 654
195 LysGlnGlnAspArgProGlyArgTrpGluHisArgThrCysArgGlnLeu 211
:::||||| :|||:|||||
655AAACCACAGCCAACTATATCAGATCATCCGGAAGGCC 693
212 LeuHisLeuLysAspGlnCys.....LeuIleSerCysCysVal... 224
||| :||| :|||:|||||:|||||:|||||:|||||:|||||
694 CTC.....TGTCGGGGTCTCTGCTGGGCTGCTCCATGCA 728
225 LeuSerProArgAlaGlyAla.....ArgGluLeuGlyG 236
:::||||| :||| :||| :|||:|||||
729 TGCTCTCCAGTGCAGCGAAGCCGAGCCATCACCAGCCAGAGCTGGTTA 778
236 LuPheHisAlaPheIleValSerAspLeuArgGluLeuGlnGlyGlnAla 252
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
779 AGAGTCATGCGGTACTCTGCTACTGGAGTCAAGAGGTTGAATTCAG... 825
253 GlyGlnCysIleLeuLeuLeuArgIleGlnAsnProTrpGlyArgGly 269
|||:||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
826 GCCATCCAGAGAGCTGATCAGACTCAGGATCCATGGGTGAAGTGA 875
269 sTrpGlnGlyLeuTrpArgGluGlyGlyGluGlyTrpSerGlnValAspA 286
|||:||||| ||| :||| :|||:|||||:|||||:|||||:|||||
876 GTGGTGGGAGCCTGGAGCGATGATGCACAGAGTGAATCACAAGACC 925
286 laAlaValAlaSerGluLeuLeuSerGlnLeuGlnGluGlyGluPheTrp 302
:::||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
926 CCGCGCGAAGAGAACTGGACAAGAAGTTGAGGATGGAGAAATCTGG 975
303 ValGluGluGluGluPheLeuArgGluPheAspGluLeuThrValGlyTy 319
:::||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
976 ATGCATCTTCAGATTTCTGAGGAGGAGTCTCTCGTGGATGATCTGCAA 1025
319 rProValThrGluAlaGlyHisLeuGlnSerLeuTrpThrGluArgLeuL 336
:::||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
1026 CTTCTCCCGGAGCTCTCTGAGTAGCAGGAGGTGCACAAATGGAACCTGG 1075
336 euCysHisThrArgAlaLeuProGlyAlaTrpValLysGlyGlnSerAla 352
::: ||| :||| :|||:|||||:|||||:|||||:|||||:|||||
1076 TCCTGTTCAAC.....GCCCACTGGACCCGGGGCTCCACAGCT 1113
353 GlyGlyCysArgAsnAsnSerGlyPhePro.....SerAsnPr 365
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1114 GGGGCTGCCAGAAC.....TACCAGCCAGCTACTGGACCAATCC 1154
365 oLysPheTrpLeuArgValSerGluProSerGlu..... 376
|||:||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
1155 CCAGTTCAAAATCCGTTTGGATGAAGTGGATGAGGAGGAGAGCA 1204

[illegible]

[illegible]

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-157-349-6

```

seq_documentation_block:
; Sequence 6, Application US/09157349
; Patent No. 6068990
; GENERAL INFORMATION:
; APPLICANT: SHINTANI, Yasushi
; APPLICANT: NISHI, Kazuori
; APPLICANT: KAWAMOTO, Tomohiro
; TITLE OF INVENTION: NOVEL PROTEINS, THEIR PRODUCTION
; TITLE OF INVENTION: AND USE
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/157,349
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/835,099
; FILING DATE:
; APPLICATION NUMBER: 97105508.2
; FILING DATE: 03-APR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Resnick, David S
; REGISTRATION NUMBER: 34,235
; REFERENCE/DOCKET NUMBER: 47342
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6440
; TELEX:
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2136 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; US-09-157-349-6

```

```
alignment_scores:
  Quality: 635.00
  Ratio: 2.096
  Percent Similarity: 59.646
  Length: 508
  Gaps: 17
  Percent Identity: 32.283
```

alignment_block:

US-09-768-877-2 x US-09-157-349-6

Align seg 1/1 to: US-09-157-349-6 from: 1 to: 2136

13 LeuPheArgAspAlaAlaPheProAlaAlaAspSerSerLeu...PheCy 28
 |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
 160 CTATTAAAGGACCGCTGAGTCTCCAGCATGTCATCAGCTTTGGCTACAA 209
 28 saspleuSerThrProLeuAlaGlnPheArgGluAspIleThrTrpArgA 45
 |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
 210 GGATCTTGGAA...CCAGGCTCTCCGCAACTCAAGGCATCATCTGGGAAGC 256
 45 rgProGlnGluileCysAlaThrProArgLeuPheProAspAspProArg 61
 |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
 257 GGCCACAGGAGTTGTGTCCAGCCCTCAGTTTATCTGTTGGTGAGCCACG 306
 62 GluGlyGlnValLysGlnGlnLeuLeuGlyAspCysThrPheLeuCysAl 78

```

307 CGCACAGACATTTGTACAGGTGCTGTAGTACCTGCTGCTTGTGCTGC 356
78 acysAlaIaleuGlnIySeraRHisIleuLeuaspGlnValIleProP 95
357 CATGGCCCTCCCTGACCCCTGATGAGAGAGCTGCTTACCGGGGTGCCCA 406
95 roGlyGlnProSerTrpAlaaspGlnIyTrpArgGlySerPheThrCys 111
407 GGCACCCAGAC.....TTCCAGGAGACTATGCGGAACTTTCACCTT 450
112 ArgIleTrpGlnPheGlyArgTrpValGlnValIleThrAspAspArgLe 128
451 CAGTCTGGCAGTACGGAGAGTGGGTGGAGTGCATTCATTCAGCACAGCT 500
128 uProCysIleuAlaGlyArgLeuCysPheSerArgCysGlnArgGlnAspV 145
501 GCCACACAGAAATGACAGCTGCTCTTCTTCTTCTTCTTCTTCTTCTTCT 550
145 alPheTrpLeuProLeuGlnIyValIyTrpAlaIyValIyValIySer 161
551 AATTCCTGAGTGCCTGCTGAGAAAGCCATGCAAGCTTAATGATGTTGT 600
162 TyrGlnHisLeuTrpAlaGlyGlnValAlaAspAlaLeuValAspLeuTh 178
601 TATGAGGCTCTGCTGAGGAGTTCACAGTGGAGGGGTTTGAGATTTCCAC 650
178 rGlyGlyLeuAlaGlyArgTrpAsnLeuIySgIyValAlaGlySerGly 195
651 AGGCGGATCTCTGAGTTTATGACTGAA..... 681
195 IyGlnGlnAspArgProGlyArgTrpGlnHisArgTrpCysArgGlnLeu 211
682 .....AAACACACAGCCAACTATATCATGATCATCCGGAAGGCC 720
212 LeuHisIleuIySaspGlnCys.....LeuIleSerCysVal.. 224
721 CTC.....TGTGGGGGCTCTGCTGGCTGCTGCTGCTGCTGCTGCTGCT 755
225 LeuSerProArgAlaGlyAla.....ArgGlnLeuGly 236
756 TGTCTCCAGTGCACCCGACCCAGCCATCACCACCCAGAGAGCTGTTA 805
236 IuPheHisAlaPheIleValSerAspLeuArgGlnLeuGlnIyGlnAla 252
806 AGAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 852
253 GlyGlnCysIleLeuLeuLeuArgIleGlnAsnProTrpGlyArgGly 269
853 GGCATCTCCAGAGAGCTGATCAGATCAGAGATCCAGATCCAGATGGA 902
269 strGlnGlyLeuTrpArgGlnIyGlyGlnIyTrpSerGlnValAspA 286
903 GTGGTGGGAGCCTGAGCATATGATCAGATCAGATGATCAGATGATCAG 952
286 IAlaValAlaSerGlnLeuLeuSerGlnLeuGlnIyGlnIyGlyPheTr 302
953 CCCGGGGAAGAGAACTGACAAAGATTGAGATGAGATGAGATGAGATG 1002
303 ValGlnGlnGlnIuPheLeuArgGlnIuPheAspIuLeuThrValGlyTy 319
1003 ATGTCACTTCAATTCGATTCGATTCGATTCGATTCGATTCGATTCGCA 1052
319 rProValThrGlnAlaGlyHisIleuGlnSerLeuTrpThrGlnArgLeu 336
1053 CCGTCCCGGAGCTCTCTGATGACAGAGAGTGCACAAATGAGACCTGG 1102
336 euCysHisThrArgAlaLeuProGlyAlaTrpValIySgIyGlnSerAla 352
1103 TCCTGTTCAC.....GGCCACTGAGACCGGGGCTCCACAGCT 1140
353 GlyGlyCysArgAsnAsnSerGlyPhePro.....SerAsnPr 365

```

```

1141 GGGGCTCCAGAAC.....TACCAGCCACAGTCTGACCAATCC 1181
365 OlySpherTrpLeuArgValSerGluProSerGlu..... 376
1182 CCAGTCAAAATCCCTTGGATGAGATGATGATGATGATGATGATGATGAT 1231
377 .....ValTrpIleAlaValLeuGlnArgSerArg 386
1232 TCGGGAACCTCCTGCTGATCAGTCTGCTGCTGCTGCTGCTGCTGCTGCT 1281
387 LeuHisAlaAlaAspTrpAlaGlyArgAlaArgAlaLeuValGlyAspSe 403
1282 .....AGTGGGGGAGACCGATGAGCAAGG 1307
403 rHisTrpSerTrpSerProAlaSerIleProGlyIyHisIyTrpGlnAlav 420
1308 CATGCTTAC.....A 1318
420 alGlyLeuHisIleuTrpIySValGlnIySarg.....Arg 431
1319 TCGGCTATGCCGCTTACAGAGTTCACAGAGCTGAGAGTACACAGGAC 1368
432 ValAsnLeuProArg.....ValIleSerMetProProValAlaGlyTh 446
1369 GCACACTTGGGCGGAGATTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1418
446 rAlaCysHisAlaTrpAspArgGlnValHisLeuArgCysGlnLeuSerP 463
1419 CAGACACTTACCTCAACCTGCGGAGGCTCTGCGGCGGCGGCGGCGGCGG 1468
463 roGlyTrpTrpLeuAlaValProSerThrPheLeuIySAspAlaProGly 479
1469 CTGGGAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1518
480 GluPheLeuLeuArgValPheSer 487
1519 GAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1542

seq_name: /cgn2_6/ptodata/2/lna/6B_COMB.seq:us-09-422-869-15
seq_documentation_block:
; Sequence 15, Application us/09422869
; Patent No. 6235481
; GENERAL INFORMATION:
; APPLICANT: POLONSKY, KENNETH S.
; APPLICANT: HORIKAWA, YUKIO
; APPLICANT: ODA, NAOMISA
; APPLICANT: COX, NANCY J.
; APPLICANT: SREENAN, SEAMUS
; APPLICANT: ZHOU, YUN-PING
; APPLICANT: OTANI, KENICHI
; APPLICANT: HANIS, CRAIG L.
; APPLICANT: BELL, GRAEME I.
; TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
; FILE REFERENCE: ARCD:307
; CURRENT FILING DATE: 1999-10-21
; EARLIER APPLICATION NUMBER: 60/134,175
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 1267
; TYPE: DNA
; ORGANISM: Human
; US-09-422-869-15

alignment_scores:
Quality: 513.50 Length: 131
Ratio: 4.75 Gaps: 3
Percent Similarity: 82.443 Percent Identity: 77.099

```

```

alignment_block:
US-09-768-877-2 x US-09-422-869-15
Align seg 1/1 to: US-09-422-869-15 from: 1 to: 1267
1 MetArgAlaGlyArgGlyAlaThrProAlaArgGluLeuPheArgAspAl 17
|||||
141 ATGGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCG 190
17 aAlaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrProL 34
|||||
191 CGCCTTCCCGCGCGGACTCTCTGCTCTTCTGGGACTTGTCTACGGCGC 240
34 euAlaGlnPheArgGluAspIleThrTrpArgArgProGlnGluIleCys 50
|||||
241 TGGCCCAAGTTCGCGAGGACATCAGCTGGAGGGCGGGCGGGCGGGCGG 290
51 AlaThrProArgLeuPheProAspAspProArgGluGlyGlnVallySGl 67
|||||
291 GCCACACCCCGGCTGTTCAGATGACCCAGGGAGGGCGGGCGGGCGGG 340
67 nGlyLeuGlyAspCysTrpPheLeuCysAlaCysAlaAlaLeuGlnL 84
|||||
341 GGGCTGCTGGGGATTGCTGGTTCCTGTGTGCTGCGCGCGCGCTGCAGA 390
84 ySSerArgHisLeuLeuAspGlnValIleProProGlyGlnProSerTrp 100
|||||
391 AGAGCAGGACCTCTCTGACAGGCTCTCTGCGCTGTGCAG..... 431
101 AlaAspGlnGluTrpArgGlySerPheThrCysArgIleTrpGlnPheG 117
|||||
432 .....CTTCTGACAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 467
117 yArgTrpValGluValThrThrAspArgLeuProCysLeu 131
|||||
468 ....TGGCTGGAG.....TTCCCATGCTG 488

```

seq_name: /cgn2_6/ptodata/2/lna/6B_COMB.seq:US-09-422-869-17

```

seq_documentation_block:
; Sequence 17, Application US/09422869
; Patent No. 6235481
; GENERAL INFORMATION:
; APPLICANT: POLONSKY, KENNETH S.
; APPLICANT: HORIKAWA, YUKIO
; APPLICANT: ODA, NAOHISA
; APPLICANT: COX, NANCY J.
; APPLICANT: SREENAN, SEAMUS
; APPLICANT: ZHOU, YUN-PING
; APPLICANT: OTANI, KENICHI
; APPLICANT: HANIS, CRAIG L.
; APPLICANT: BELL, GRAEME I.
; TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
; FILE REFERENCE: ARCD:307
; CURRENT APPLICATION NUMBER: US/09/422,869
; EARLIER FILING DATE: 1999-10-21
; EARLIER FILING DATE: 1999-05-13
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 864
; TYPE: DNA
; ORGANISM: Human
US-09-422-869-17

```

```

alignment_scores:
Quality: 472.00 Length: 92
Ratio: 5.130 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:

```

```

US-09-768-877-2 x US-09-422-869-17
Align seg 1/1 to: US-09-422-869-17 from: 1 to: 864
581 GlnValProGluGlyArgSerGlnAspAlaProProLeuLeuGlu 597
|||||
180 CAGGTCCAGAGGGTGGAGAGCCAGGAGCCACCCCACTGCTGCTGCA 229
597 nGluProLeuLeuSerCysValProHisArgTyrAlaGlnGluValSerA 614
|||||
230 GGAGCCGCTGCTGAGCTGGCTGCCACATCGTCCGCGGAGGAGTGA 279
614 rGLeuCysLeuLeuProAlaGlyThrTyrLysValValProSerThrTyr 630
|||||
280 GGCTCTGCCCTCTGCTGCGTGCAGGACCTACAGGTTGTGGCCCTCCAC 329
631 LeuProAspThrGluGlyAlaPheThrValThrIleAlaThrArgIleAs 647
|||||
330 CTCCGCGACACAGAGGGGCGCTTCACAGTGACCATCGCAACAGGATTGA 379
647 pArgProSerIleHisSerGlnGluMetLeuGlyGlnPheLeuGlnGlu 664
|||||
380 CAGGCCATCCATTTCACAGCAGGAGATGCTGGGCCAGTTCTCTCCAGAG 429
664 alSerValMetAlaValMetLysThr 672
|||||
430 TCTCCGTCATGCGCAGTGATGATAAAC 455

```

seq_name: /cgn2_6/ptodata/2/lna/5B_COMB.seq:US-08-928-692-16

```

seq_documentation_block:
; Sequence 16, Application US/08928692
; Patent No. 5958727
; GENERAL INFORMATION:
; APPLICANT: Brody, Howard
; APPLICANT: Taver, Deborah S.
; APPLICANT: Lamsa, Michael
; APPLICANT: Hansen, Kim
; TITLE OF INVENTION: Methods for Modifying the Production of
; a Polypeptide
; NUMBER OF SEQUENCES: 80
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5958727o No. 5958727disk of No. 5958727th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,692
; FILING DATE: 12-SEPT-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4944.200-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4700 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-928-692-16

```

alignment_scores:

Quality: 263.00 Length: 677
 Ratio: 0.843 Gaps: 32
 Percent Similarity: 46.086 Percent Identity: 21.861

alignment block:

US-09-768-877-2 x US-08-928-692-16

Align seg 1/1 to: US-08-928-692-16 from: 1 to: 4700

```

43 TrpArgProGlnGluIleCysAlaThr.....Pr 53
|||||
1944 TGGAAACGACCTTCGGAGCTTCGGCAAGACAGACGAGATGTGACAC 1993
53 oArgLeuPhePro.....AspAspProArgGluGluValIysGlnG 68
|||||
1994 AAGGTGGTTCCTGTGATGTCTGTCCAGAAAGACAGATCTGTCCAGG 2043
68 IyLeuLeuGlyAspCysTirPhe.....LeuCysAlaCysAlaAla 81
|||||
2044 ATATGTGTACGGACTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 2093
82 ...LeuGlnIlySerArgHisLeuLeuAspGlnValIleProGluG 97
|||||
2094 ATGCTGAACCGCGCCAGTGTACTGTAGAGATGTATCCCTTCGGCTG 2143
97 nProSerTrp..... 100
|||||
2144 ACCGTGATGTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 2193
101 .....AlaAspGlnGluTyrArg..GlySerPheThrC 111
|||||
2194 ATACCTGACCGCGGGAGCTGTACAGCTTCACCGTCAGCGCAAGTATAT 2243
111 ysArgIleTrpGlnPheGlyArgTyrValGluValThrThrAspArg 127
|||||
2244 TTCGGCTTTATTTCAATGGGTCTTCGGAAAGTCACTTATGACGACGT 2293
128 LeuProCysLeuAlaGlyArgLeuCysPheSerArgCysGlnArgGlu 143
|||||
2294 TTGGCATGCG.....TCTAAGACATCAAGATCACT 2322
144 .....AspValPheTrpLeuProLeuLeuG 152
|||||
2323 CCACGTGATCGACCGGAAATCCCAATTTCTTTGGCGCGCGCTGTAG 2372
152 IuLyValTyrAlaIysValHisGlySerTyrGlnHisLeuThrAlaG 168
|||||
2373 AGAAGCGGTATTTGAATTTGCCGAGAGCTATGATTTTCCGGAAGCAAT 2422
169 GlnValAlaAspAlaLeuValAspLeuThrGlyGlyLeuAlaGluArg 185
|||||
2423 TCCGGAGACAGATCTGTGGTG...CTGACAGTTGATTTCCGAGCAAGT 2469
185 PAsnLeuGlyValAlaGlySerGlyGlnGlnAspArgProGlyAla 202
|||||
2470 CTTTCTCATATATGACGATGTGACGGCGACCG.....C 2504
202 rGTrpGlnHisArgThrCysArgGlnLeuLeuHisLeuLysAspGlnC 218
|||||
2505 TCTGGAG...CGACTTACAGATCC...TTTCAACCAAGAGATGTCTC 2548
219 Leu...IleSerCysCysValLeuSerProArgAlaGlyAlaArgGlu 234
|||||
2549 TTGACTATAGTACCGGTGACTCACTGAGAGGAA...CAAGAAGAACT 2595
234 uGly.....GluPheHisAlaPheIleValSerAspLeuArgGluG 249
|||||
2596 AGGCTCTGTGTGACGATGTATGTCTGTATCTGTATGAAGAAATCTA 2645
249 IngLyGlnAlaGlyGlnCysIleLeuLeuLeuArgGlyGlnAspPro 265
|||||

```

```

2646 AAGTGCACGA.....CAATTACTCGTGAAGAAACCTTGG 2680
266 GlyArgArgCysTrpGlnGlyLeuTrpArgGluGlyGlyGlyTyrP 282
|||||
2681 GCT.....GGACGAGATCTGCCCCCGCGACAAATGGAAG 2715
282 rGlnValAspAlaAlaValAlaSerGluLeuLeuSerGlnLeuGlnG 299
|||||
2716 CCTCTGTGATCCAGAGATTTACCCATTAACCGCCCTATTTGACCCG 2765
299 IyLeuPheTrpValGluGluGluGluPheLeuArgGluPheAspGlu 315
|||||
2766 GTACCTTTTGATGATGTGGAAAGCTGTTCATCAATTTGAAACCTC 2815
316 ThrValGlyTyrProValThrGluAlaGlyHisLeuGlnSerLeu 332
|||||
2816 TATTTGAAT.....TGAA 2829
332 rGluArgLeuLeuCysHisThrArgAlaLeuProGlyAlaTrp.... 347
|||||
2830 CCGTGAATTTTCAATACCGGCAAGACGTCCACTTACGTGGAGACTCA 2879
347 aLysGlyGlnSerAlaGlyGlyCysArgAsnAsnSerGlyPheProSer 363
|||||
2880 ACAACGGAGAGGTGTACCGGCTGT.....TTGTGAT 2914
364 AsnProLysPheTrpLeuArgValSerGluProSerGluValTyrLeu 380
|||||
2915 AACCGGAGTGTGCAAGTGTCAACCGAAGCGGTGGATGTCTGTGTTA.. 2962
380 aValLeuGlnArgSerArgLeuHisAlaAlaAspTrpAlaGlyAla 397
2962 ..... 2962
397 rGAlaLeuValGlyAspSerHisThrSerTrpSerProAlaSerIlePro 413
|||||
2963 .....CTTCTAGCAAGCATTTCAAGACAAACAGGCGACCGAAGACT 3007
414 GlyLysHisTyrGlnAla.....ValGlyLeuHis 424
|||||
3008 CTTGACGAAATACCAAGCAAGTAGAGATCGGCTTTATTAACATATAGT 3057
424 uTrpLysValGluLysArgArgValAsnLeuProArgValLeuSerMet 441
|||||
3058 CTTTAACCGAATGGCAACAGGCTCTCTTTGAGT..... 3091
441 rProValAlaGlyThrAlaCysHisAlaTyrAspArgGluValHis... 456
3092 .....GATGGGCGCTTACATCGT 3109
457 .....LeuArgCysGluLeuSerPr 463
3110 GGGCCCTGTGTGGATTCCTTAATACGCTCATGAGCTTAGAGATGCC 3159
463 O.....GlyTyrTyrLeuAlaValProSerThrPheLeuLysAspAla 478
|||||
3160 CAGAAACAAATACACAGATCGTGTCTCCGACAAATCACTGCCATCTTGA 3209
478 rGlyGlnPheLeuLeuArgValPheSerThrGlyArgValSerLeuSer 494
|||||
3210 ATCAAACTTTACTTTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 3259
495 AlaIleArg.....AlaValAlaLysAsnThrThrProGlyAla 507
|||||
3260 AAGCCCAAGATTAATATACATGTGTCTAGAG..... 3292
507 aAlaLeuProAlaGlyGluTrpGlyThrValGlnLeuArgGlySerTrp 524
3293 .....ATTCAGAGGCTTTGG 3307
524 rValGlyGlnThrAlaGlyLysArgAsnPheAlaSerTyrProThr 540
|||||
3308 ..ACACCTTGACGCGAGGTGGATGCGCAATCTTCTCATATTCACCTC 3355

```

```

541 AsnProCysPheProPheSerValProGluGlyProGlyPro..... 554
||||| ||| :::::::::: ::
3356 AACCCCAATTAGTTGGAGATAGAGAAATGACACAGATGTTTCACTCCT 3405
555 ....AtgCysValArgIleThrLeuHisGlnHisCysArgProSerAspT 570
::: ||| ::
3406 GCTGGATGCCCAACACGGAACCTCCGACCCCATGTTAAG..... 3445
570 hrGluPheHisProIleGlyPheHisIlePheGlnValProGluGlyGly 586
||| ||| :::::::::: ::
3446 ..TTATTCGTGTCCTAATGGAATCGTGTCTCGGAGTA..... 3481
587 ArgSerGlnAspAlaProProLeuLeuLeuGlnGluProLeuLeuSerCy 603
||||| ::
3482 CGCAGTCGGACATA.....ATCGCTGATAG 3507
603 sValProHisArgTyrAlaGlnGluValSerArgLeuCysLeuLeuProA 620
::: ||| :::::::::: ::
3508 TGGTGACTATCGCGGTGGTGGCTCCCTGTGGAAGAAAGGCTCTGGAAC 3557
620 laGlyThrTyrLysValValProSerThrTyrLeuProAspThrGluGly 636
||: ||| :::::::::: ::
3558 CGGGCTCATATACATCTGTCTTCCACATTCGCGCGGATCAACTTGGC 3607
637 AlaPheThrValThrIleAlaThrArgIle 646
||||| :::::::::: ::
3608 CGATTCACGCTCTGGGTATCTCTTCTAGTT 3637

```


|||||
274 GCCACACCCCGCTGTTTCCAGTAGCCACCGGAAGGCGAGGTGAAGCA 323
67 nGlyLeuLeuGlyAspCysTrpPheLeuCysAlaCysAlaLeuGlnL 84
|||||
324 GGGGCTGCTGGGGGATGCTGTTCTGTGCTGCTGCGCGCGCTGCAGA 373
84 ySerArgHisLeuLeuAspGlnValIleProProGlyGlnProSerTrp 100
|||||
374 AGAGCAGGCACCTCTCGGACCAAGTCATCTCTCGGACAGCGAGCTGG 423
101 AlaAspGlnGluTrpArgGlySerPheTrpCysArgGlyTrpGlnPhe 117
|||||
424 GCGCACCAGGAGTACCGGGGCTCTTCACTCTGCTGCTGCTGCTGCTG 473
117 yArgTrpValGluValThrAspAspArgLeuProCysLeuAlaGlyA 134
|||||
474 AGCTGGGTGGAGTGACACAGATGACCGCTGCGCTGCTGCTGCTGCTG 523
134 rGLeuCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
|||||
524 GACTCTGTTCTCCCGCTGCCAGAGGAGGATGTGTTCTGCTGCTGCTG 573
151 LeuGluLeuValTyAlaLysValHisGlySerTyArgGluHisLeuTrp 167
|||||
574 CTGGAAAGGCTACGCCAAGGTCATCGGCTCTACGAGCACCTGTGGGC 623
167 aGlyGlnValAlaAspAlaLeuValAspLeuThrGlyGlyLeuAlaGlu 183
|||||
624 CCGGAGGTGGGGATGCTGCTGGGTGGACCTGACCGCGGCGCTGCAGAA 673
184 ArgTrpAsnLeuLysGlyValAlaGlySerGlyGlnGlnAspArgPr 200
|||||
674 AGATGAACCTGAAGGGGTASAGGAGGAGGAGGAGGAGGAGGAGGAGG 723
200 cGlyArgTrpGluHisArgThrCysArgGlnLeuLeuHisLeuLysAspG 217
|||||
724 AGCGCGCTGGGAGCAGAGCTGTGCGGAGCTGCTCCACCTGAAGGACC 773
217 InCysLeuIleSerCysValLeuSerProArgAlaGlyAlaArgGlu 233
|||||
774 AGTGCTGATCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 823
234 LeuGlyGluPheHisAlaPheIleValSerAspLeuArgGluGlnG 250
|||||
824 CTGGGGGAGTTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 873
250 yGlnAlaGlyGlnCysIleLeuLeuLeuArgIle 261
|||||
874 TCAGCGGGGCGAGTGCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 907

seq_name: gb_est2:B1916704

seq_documentation_block:
LOCUS B1916704 887 bp mRNA linear EST 16-OCT-2001
DEFINITION 603178679F1 NIH_MGC_121 Homo sapiens cDNA clone IMAGE:5242963 5',
mRNA sequence.
ACCESSION B1916704
VERSION B1916704.1 GI:16180666
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 887)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11612 row: n column: 20
High quality sequence stop: 809.

FEATURES

source
1..887
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5242963"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: EcoRV (destroyed); RNA source anonymous pool of 3
fetal brains, female age 20 weeks, female age 24 weeks,
and male age 26 weeks. Library is oligo-dT primed and
directionally cloned (EcoRV site is destroyed upon
cloning). Average insert size 1.7 kb, insert size range
0.7-3.5 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(Invitrogen). Research Genetics tracking code 017. Note:
this is a NIH_MGC Library."
BASE COUNT 163 a 260 c 310 g 154 t
ORIGIN

alignment_scores:

Quality: 1304.00 Length: 299
Ratio: 4.559 Gaps: 8
Percent Similarity: 95.652 Percent Identity: 92.308

alignment_block:

US-09-768-877-2 x B1916704

Align seg 1/1 to: B1916704 from: 1 to: 887

211 LeuLeuHisLeuLysAspGlnCysLeuIleSerCysValLeuSerPr 227
|||||
2 CTGCTCCACCTGAAGGACCACTGCTGATCAGCTGCTGCTGCTGCTGCTG 51
227 cArgAlaGlyAlaArgGluLeuGlyGluPheHisAlaPheIleValSerA 244
|||||
52 CAGAGCAGGTGCGCGGAGCTGGGGAGTCCATGCCCTTCATTGCTCGG 101
244 sPLeuArgGluLeuGlnGlyGlnAlaGlyGlnCysIleLeuLeuLeuArg 260
|||||
102 ACCTGCGGGAGCTCCAGGGTCAGGCGGCCAGTGCATCTGCTGCTGCGG 151
261 IleGlnAsnProTrpGlyArgArgCysTrpGlnGlyLeuTrpArgGluG 277
152 ATCCAGAACCCCTGGGCGCTGGCGGCTGCTGCGAGGGGCTCTGGAGAGG 201
277 lyGlyGluGlyTrpSerGlnValAspAlaAlaValAlaSerGluLeuLeu 293
|||||
202 GGGGTGAAGGTGAGCAGCAGGTAGATGCGCGGTAGCATCTGAGCTCGTG 251
294 SerGlnLeuGlnGluGlyGluPheTrpValGluGluGluPheLeuArg 310
252 TCCAGGCTCCAGGAAGGGGAGTCTCTGGGTGGAGGAGGAGGAGTTCCTC 301
310 gGluPheAspGluLeuThrValGlyTyProValThrGluAlaGlyHisL 327
302 GGAGTTTGACGAGCTACCGTTGGCTACCGGCTCAGGAGGCGCGGCCACC 351
327 euGlnSerLeuTyThrGluArgLeuLeuCysHisThrArgAlaLeuPro 343
352 TGCAGAGCCTCTACACAGAGAGGCTGCTCTGCCATACGCGGGCGCTGCT 401
344 GlyAlaTrpValLysGlyGlnSerAlaGlyCysArgAsnSerG 360
402 GGGGCCCTGGGTCAAGGGGCCAGTTCAGCAGGAGGCTGCCGGAACAACAGCGG 451


```

360 yPheProSerAsnProLysPheTrpLeuArgValSerGluProSerGluVal 377
452 CTTTCCACCAACCCCAATTTCTGCTGGGGCTCAGAACCGATGAGG 501
377 aTyrTLeaValLeuGlnArgSerArgLeuHisAlaAlaAspTrpAla 393
502 TGTACATTCCTGCTCGACAGATCCAGCTGACCGCGGAGCTGGGGA 551
394 GtArgAlaArgAlaLeuValGlyAspSerHisThrSerTrpSer. ProA 410
552 GGGCCGGCC. CGGGCACTGGTGGGTGACACTGACTTCTGAGGCCCCAG 600
410 laSerTLeProGlyLysHisTyrGlnAlaValGlyLeuHisLeuTrpLys 426
601 CGAGCATCCCGGGCAAGCACTACAGAGCTGTGGGTCTGCACCTGTGGAAG 650
427 .ValGluLysArgArgVal. AsnLeuProArg. ValLeuSerMetProPr 442
651 GGTAGAGAAGCGCGGGTCAATCTGCCTAGGGGCTCTTCATGCCCC 700
442 oValAlaGlyThrAlaCysHisAlaTyrAspArgGluValHisLeuArgC 459
701 CGTGGCTGGCACCGGTGCCAATCTGCATACGACCGGAGGTCCACTGGCT 750
459 ySGluLeuSerProGlyTyrTyrLeuAlaValProSerThrPheLeuLys 475
751 GTGACCTCTCAG. GGTACTACTCTGAGCTGTCCAG. AACTTCTGGAAG 798
476 AspAlaProGlyGluPheLeuLeuArgVal. PheSerThrGlyArg. Val 491
799 GACGGCGCAGGGAAGTCTCTGCTCGAGCTTACTACTGCGGAGAAATGC 848
492 SerLeuSerAlaIleArgAlaValAlaLys 501
849 TCCTTAGGGCCATCCAGCAAGTGCAGA 879

seq_name: gb_est2:BG475966

seq documentation block:
LOCUS BG475966 842 bp mRNA linear EST 21-MAR-2001
DEFINITION 602521077F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:4639579 5',
mRNA sequence.
ACCESSION BG475966
VERSION BG475966.1 GI:13408245
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 842)
REFERENCE NIH-MGC http://mgi.ncl.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC/DCPD/DPF
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LITCMI403 row: 1 column: 20
High quality sequence stop: 760.
Location/Qualifiers
1..842
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image:4639579"
/clone_lib="NIH_MGC_20"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pORF7; Site_1: XhoI; Site_2:

```

```

ECORI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
BASE COUNT 164 a 238 c 307 g 133 t
ORIGIN

alignment_scores:
Quality: 1241.50 Length: 282
Ratio: 4.721 Gaps: 11
Percent Similarity: 93.262 Percent Identity: 91.135

alignment_block:
us-09-768-877-2 x BG475966

Align seg 1/1 to: BG475966 from: 1 to: 842

193 SerGlyGlyGlnGlnAspArgProGlyArgTyrGlnHisArgThrCysAr 209
|||||
2 AGCGGAGGCCAGACAGACAGGCGCGCTGGGAGACAGACTTGTGCG 51
gGlnLeuLeuHisLeuLysAspGlnCysLeuIleSerCysValLeu 226
|||||
52 GCAGCTGCTCCACCTGAAGACCACTGTCTGATCACTGCTGCTGCTCA 101
erProArgAlaGlyAlaArgGluLeuGlyGluPheHisAlaPheIleVal 242
|||||
102 GCGCCAGACAGAGTCCCGGAGGAGTGGGGAGTTCATTCCTTATTTTC 151
SerAspLeuArgGluLeuGlnGlyGlnAlaGlyGlnCysIleLeuLeu 259
|||||
152 TCGGACCTGCGGGAGCTCCAGGCTCAGGGCGCCAGTCCATCTGCTCT 201
uArgIleGlnAsnProTyrGlyArgCysTyrPdlGlnIleLeuTrpArg 276
|||||
202 GCGGATCCAGAACCCCTGGGGCGCGGCTGCTGCGAGGGCTCTGGAAG 251
276 IugIyGlyGlyGlyTyrSerGlnValAspAla. AlavalalaSerGluLe 292
|||||
252 AGGGGGGGAAGGCTGAGCCAGGTAGATCAAGCGGTAGCATCTGACT 301
uLeuSerGlnLeuGlnGlyGluGlyGluPheTrpValGluGluGluPheL 309
|||||
302 CTTGTCACAGCTCCAGGAAGGGAGTCTGGTGGAGAGAGAGAGTTC 351
euArgGluPheAspGluLeuThrValGlyTyrProValThrGluAlaGly 325
|||||
352 TCAGGAGGTGGAGAGAGCTACCTAGGCTACCGGTCACGAGGCCGGC 401
326 HisLeuGln. SerLeuTyrThrGluArgLeuLeuCysHisThrArgAla 342
|||||
402 CACCTGCAGAGACCTCTACACAGAGAGCTCTGTCATATCGGGGGGC 451
eupProGlyAlaTrpValGlyGlnSer. AlaGlyGly. CysArgAsnA 358
|||||
452 TGCCTGGGGCTGGGTCAAGGGCCAGTCCAGCAGAGAGCTGACCGGACA 501
358 snSerGlyPheProSer. AsnProLysPheTrpLeuArgValSerGlu 374
|||||
502 ACAGCGGCTTACCCCAAGAACCCCAATTTCTGCTCGGGCTCTCAAC 551
374 roSerGluValTyrTLeaValLeuGlnArgSerArgLeuHisAlaAla 390
|||||
552 CGAGTGGGTGTACATAGCTGCTGCAAGATCCAGGCTGACGGCGCA 601
391 .AspTrpAlaGlyArg. AlaArgAlaLeuValGlyAspSerHisThrSer 406
|||||
602 GGACTGGGAGCGCGGGCGCGGCTGCTGCTGCTGCTGCTGCTGCTG 651

```


seq_name: gb_est2:B1668734

seq_documentation_block:
 LOCUS B1668734 738 bp mRNA linear EST 12-SEP-2001
 DEFINITION 603293173F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:5312178 5',
 mRNA sequence.
 ACCESSION B1668734
 VERSION B1668734.1 GI:15582967
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 738)
 AUTHORS NIH-MGC http://mgs.ncl.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: c9abs-r@mail.nih.gov
 Tissue Procurement: Miklos Palokovits, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki
 Toshiyuki and Piero Carninci (RIKEN)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: LLAM1191 row: b column: 19
 High quality sequence stop: 738.
 Location/Qualifiers
 1..738
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5312178"
 /clone_1lb="NIH_MGC_96"
 /tissue_type="hypothalamus"
 /lab_host="DH10B"
 /note="Organ: brain; Vector: pBluescript (modified
 Bluescript KS+); Site: 1: BamHI; Site 2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTAA-3',
 size-selected for average insert size 2.3 kb and
 normalized to ROT 5. This is a primary library enriched
 for full-length clones and constructed using the
 Cap-trapper method (Carninci, in preparation). Library
 constructed by M. Brownstein (NIH/NHGRI, National
 Institutes of Health). Note: this is a NIH-MGC Library."
 BASE COUNT 120 a 255 c 237 g 126 t
 ORIGIN

alignment_scores:
 Quality: 1163.00 Length: 243
 Ratio: 5.035 Gaps: 0
 Percent Similarity: 95.062 Percent Identity: 93.004

alignment_block:
 US-09-768-877-2 x B1668734 ..

Align seg 1/1 to: B1668734 from: 1 to: 738

369 leuArGValSerGIuProSerGIuValTyrTleAlaValleuGlnArgSe 385
 :::
 6 GTGGGGGTCTCAGAACGAGTGTACTTCCCTCCGCGAGATC 55
 385 rArgLeuHisAlaAlaAspTrpAlaGlyArgAlaArgAlaLeuValGly 402
 :::
 56 CAGGCTCAGCGCGGCGACTGGCGACCGCGCCCGGCGCACTGGTGG 105
 402 spSerHisThrSerTrpSerProIleSerIleProGlyLysHisTyrGln 418
 :::
 106 ACACTCTACTCTGTGGAGCCACGAGCAATCCCGGCGAAGCACTACAG 155

seq_name: gb_est1:AL529373

seq_documentation_block:
 LOCUS AL529373 787 bp mRNA linear EST 13-FEB-2001
 DEFINITION AL529373 LTR_NFL001.NBC4 Homo sapiens cDNA clone CSDD0006VAL5 5
 prime, mRNA sequence.
 ACCESSION AL529373
 VERSION AL529373.1 GI:12792866
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 787)
 AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
 Location/Qualifiers
 1..787
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

419 AlaValGlyLeuHisLeuTrpIleValGlyLysArgAlaValLeuLeuPr 435
 :::
 156 GCTGTGGGTCTGCACCTGTGGAAGTAGAAGCGCGGCTCAATCTCC 205
 435 oArGValLeuSerMetProProValAlaGlyThrAlaCysHisAlaTyr 452
 :::
 206 TAGGCTCTGTCATGATGCCCGCTGGTGGCAGCCGCTGCATGATC.G 254
 452 sPArgGlyValHisLeuArgCysGluLeuSerProGlyTyrTyrLeuAla 468
 :::
 255 ACCGGAGGTCCACCTGCTGTGAGCTTCACCGGCTACTACTGGCT 304
 469 ValProSerThrPheLeuLysAspAlaProGlyLysPheLeuAlaYva 485
 :::
 305 GTCCCGACCTCTCTTAAGAGCGCCAGGAG.TTCCTGCTCCGAGT 353
 485 lPheSerThrGlyArgValSerLeuSerAlaIleArgAlaValAlaTys 502
 :::
 354 CTTCCTACCGCGGCGAGTCTCCCTTAGCGCATCAGGCACTGGCCAGA 403
 502 snrThrProGlyAlaAlaLeuProAlaGlyLysTrpGlyThrValGln 518
 :::
 404 ACACCAACCCCGGCGAGCCCTGCTGGCGGAGAGTGGGAGCGTGAG 453
 519 leuArGlySerTrpArgValGlyGlnThrAlaGlyLysArgAsnph 535
 :::
 454 CTACGGGTTCTTGAGAGTCCGCGCAGACGCGGGGCGCACAGAACTT 503
 535 eAlaSerTrpProThrAsnProCysPheProPheSerValProGluGlyP 552
 :::
 504 TGCCCTCAACCCCAACCAACCTGCTTCCCTTCGCGGCCCCAGAGGCC 553
 552 roGlyProArgCysValArgIleThrLeuHisGlnHisCysArgProSer 568
 :::
 554 CTGGCCCCCGGTGGGTCCGCACTGTGCATCGCGCACTGTGGCCCCAGT 603
 569 AspThrGluPheHisProIleGlyPheHisIlePheGlnValProGluG 585
 :::
 604 GACACCGAGTTCACACCCATGCGCTTCATATCTTCACAGGCCAGAGGG 653
 585 yGlyArgSerGlnAspAlaProProLeuLeuGlnGluProLeuLeuS 602
 :::
 654 TGAAGAGGAGCAGAGCAGACCCCACTGCTGTGCAGAGACCGTGTGCA 703
 602 erCysValProHisArgTyrAlaGlnGlu 611
 :::
 704 GCTGCTGCCACATCGCTACGCCAGGAG 732

485 alpHeserThrGlyArgValSerLeuSerAlaIleArgAlaValAlaLys 501
 101 TCTTCTTACCGGGGAGTCTCCCTTAGCCATCAGGCGACATGGCCAG 150
 502 AsnThrThrProGlyAlaIleAlaLeuProAlaGlyIleThrGlyThrValG1 518
 151 AACACCGCCCGGGGACACCTGCTGCGGGGAGTGGGGACCGTGA 200
 518 nLeuArgGlySerThrPArgValGlyGlnThrAlaGlySerArgAsp 535
 201 GCTACGGGGTCTTGAGAGTGGCCAGACGCGGGGGGACACAGAACT 250
 535 hAlaSerThrProThrAsnProCysPheProPhaSerValProGlyGly 551
 251 TTGCTCATACCCACCAACCCCTCTCCCTCTCCGCTCCGAGGCG 300
 552 ProGlyProArgCysValArgIleThrLeuHisGlnHisCysArgProse 568
 301 CTTGGCCCGCGCTGCTGCGGACATCTGCAATCAGCACTGCGGCCAG 350
 568 rAspThrGluPheHisProIleGlyPheHisIlePheGlnValProGlyG 585
 351 TGACACCGAGTTCACACCCATCGGCTTCATATCTCCAGTCCAGAG 400
 585 LysGlyArgSerGlnAspAlaProProLeuLeuGlnGluProLeuLeu 601
 401 GTGAGAGGAGCGACGACCCACCCACATGCTGTCAGAGCCGCTGCTG 450
 602 SerCysValProHisArgTyrAlaGlnGlnValSerArgLeuCysLeu 618
 451 AGCTCGGCGCCACATCGCTC GCCAGAGAGTGACCGGCTGCTGCTCT 499
 618 uProAlaGlyThrTyrLysValValProSerThrTyrLeuProAspThrG 635
 500 GCTGCGGGGACACTACAGAGTGTGCTCCCTCACACTACCTGCGGAG 549
 635 LysGlyAlaPheThrValThrIleAlaThrArgGlyLeuAspArgProSerIle 651
 550 AGGGGGCTTCACAGTGCATCGCAACCCAGAGATTGACAGGCCATCCAT 599
 652 HisSerGlnGluMetLeuGlyGlnPheLeuGlnGlnValSerValMetAl 668
 600 CACAG CAGAGATGCTGCGGCACTTCTCCAGAGGCTCTCCGTCATGCG 648
 668 aValMetLysThr 672
 649 AGTCATGAAAAC 661

seq_name: gb_est2:BF310408

seq_documentation_block:

LOCUS BF310408 942 bp mRNA linear EST 21-NOV-2000
 DEFINITION 601895061F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:4124630 5',
 mRNA sequence.
 ACCESSION BF310408
 VERSION BF310408.1 GI:11257971
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 942)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov

FEATURES
 Source
 Plate: LCM1011 row: 1 column: 15
 High quality sequence stop: 718.
 Location/Qualifiers
 1..942

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4124630"
 /clone_lib="NIH_MGC_19"
 /tissue_type="neuroblastoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: brain; Vector: pORF7; Site:1: XhoI; Site:2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCAAGAG(G). Library constructed by Ling Hong
 in the laboratory of Gerald W. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH_MGC library."

BASE COUNT 193 a 250 c 347 g 152 t
 ORIGIN

alignment_scores:
 Quality: 1106.00 Length: 309
 Ratio: 4.337 Gaps: 4
 Percent Similarity: 82.524 Percent Identity: 77.346

alignment_block:

US-09-768-877-2 x BF310408

Align seg 1/1 to: BF310408 from: 1 to: 942

216 AspGlnCysLeuIleSerCysValLeuSerProAlaGlyAlaAr 232
 3 GACCACTGTGATCAGCTGCTGCTGCTGCAACCCACAGAGAGTGGCCG 52
 232 gGluLeuGlyLupPheHisAlaPheIleValSerAspLeuArgGluLeu 249
 53 GGAGCTGGGGAGATTCCATGCCCTTCATGTCTCGGACTGCGGAGCTCC 102
 249 LngGlyAlaGlyGlnCysIleLeuLeuLeuArgGlyLeuAsnProTrp 265
 103 AGGCTCAGCGGGGCGCAGTGCATCTGCTGCGGATTCACAAACCCCTG 152
 266 GlyArgArgCysTrpGlnGlyLeuTrpArgGluGlyGlyGlyTrpSe 282
 153 GCGCGCGCTGCTGCGAGGGCTCTGCGAGAGAGGGGCTGAAAGGTGAG 202
 282 rGlnValAspAlaIleAlaValAspGluLeuLeuSerGlnLeuGlnGlu 299
 203 CCAGGTGATGACAGCGGAGCATCTGAGCTCTCCACGTCACAGAG 252
 299 LysGluPheTrpValGluGluGluGluPheLeuArgGluPheAspGluLeu 315
 253 GCGAGTCTGCGTGGAGAGAGAGAGAGTCCCTCAGAGCACTTTCAGACCTC 302
 316 ThrValGlyTyrProValThrGlnAlaGlyHisLeuGlnSerLeuTyrTh 332
 303 ACCCTTGCTACCCCGGTACAGAGAGCGCG CACCTGCAGACCTCTACAC 351
 332 rGluArgLeuLeuCysHisThrArgAlaLeuProGlyAlaTrpValLys 349
 352 AGAGAGCTGCTGCTGCAATACGCGGGGCTGCTGCGGCTGAGG 401
 349 LysGlnSerAlaGlyGlyCysArgAsnAsnSerGlyPheProSerAspPro 365
 402 GCCAGTCAGAGAGAGGCTGCGCGAACAAGCGGCTTCCACAGACCCC 451
 366 LysThrPheLeuArgValSerGluProSerGluValTyrIleAlaValLe 382
 452 AAATTCGCTGCGGCTGCTGAGAACCGAGTGAAGGTGTACATTCGCGCT 501
 382 uGlnArgSerArgLeuHisAlaIleAspTrpAlaGlyArgAlaArgAlaL 399

|||||
502 GCAGAGATCCAGGCTGCACGGCGG . GACTTGGCAGCGG . GCCCGGCCAC 549
|||||
399 euValGlyAspSerHisThrSerTrpSerProAlaSerIleProGlyLys 415
|||||
550 TGGTGGGTGCACATCATCTCGTGAGCCACAGCAGCATCCCGGCCAAG 599
|||||
416 HisTyrGlnAlaValGlyLeuHisLeuTrpLysValGluLysArgVa 432
|||||
600 CACTACCAGGCTGTGGGTTCACCTCTGGAAGGTAGAGAAGCGGG . GT 648
|||||
432 lasLeuProArgValLeuSerMetProProValalaglythrAlaCysH 449
|||||
649 CAATCTGCTAGGTCCTGTATGCCCGCGGTGGACCGCTGCCATGC . 697
|||||
449 isaLatyrAspArgGluValHisLeuArgCysGluLeuSerProGlyTy 465
|||||
698 ATACGAGGGAGTCACCTGCTGGGAAGCTCTTACCGGGTTAT 739
|||||
466 Tyr LeuAlaValProSerThrPheLeuLysAspAlapr 478
|||||
740 ACTGGTGGCCAGACCTCTGGAGAGCGCAGGAGTCTCTGGCGGAGTCCTTA 789
oGlyGluPheLeuLeuArg
|||||
478 484
790 CGGGGATTCTTAGAAAGAAGGATCGAANAACACCGGGGAGCGGCTCGGG 839
485 ValPheSerThrGlyArgValSerLeuSerAlaIleArg 497
:::|||||
840 TTGGGGAGCACACGCTCGGAACGCCCAAGGGGGA..... 874
498 AlaValalalysAsnThrThrProGly 506
::: |||||
875 AGAATTCGCCACAGCAGGACACCGGGA 901

```

seq_name: gb_htc:AK002548

seq_documentation_block:
LOCUS      AK002548                      1137 bp      mRNA      linear      HTC 19-JAN-2002
DEFINITION Mus musculus adult male kidney cDNA, RIKEN full-length enriched
            library, clone:061001119:caipain 10, full insert sequence.
ACCESSION AK002548.1 GI:12832609
VERSION   Mus musculus (strain:C57BL/6J) adult male kidney cDNA to mRNA,
KEYWORDS  HTC; CAP trapper.
SOURCE    clone.lib:RIKEN full-length enriched mouse cDNA library
            clone:061001119.

ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  1 (sites)
AUTHORS   Carninci,P. and Hayashizaki,Y.
TITLE     High-efficiency full-length cDNA cloning
JOURNAL   Meth. Enzymol. 303, 19-44 (1999),
MEDLINE   99279253
PUBMED    10349636

REFERENCE  2 (sites)
AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE   204939374
PUBMED    11042159

REFERENCE  3 (sites)
AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishino,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujimoto,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE     RIKEN integrated sequence analysis (RISA) system--384-format

```

sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
MEDLINE
PUBMED
11076861
4 (sites)
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001),
5 (bases 1 to 1137)
Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A.,
Arakawa, T., Baldarelli, R., Bono, H., Brownstein, M., Bulc, C.,
Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T.,
Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hiraoka, T., Hori, F.,
Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S.,
Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K.,
Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J.,
Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D.,
Schraml, L., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,
Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yamanaka, I.,
Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and
Hayashizaki, Y.
Direct Submission
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp,
URL: <http://genome.gsc.riken.go.jp/>, tel: 81-45-503-9222,
Fax: 81-45-503-9216)
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for
further details.

```

cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues. First strand cDNA was primed with a primer
15' GAGAGAGAGCGCCCAACTCGAGTGTGTTTTTTTTTTTN 3'], cDNA was
and subsequently enriched for full-length by cap-trapper. cDNA went
through one round of normalization to Rot - 5.0. Second strand cDNA
was prepared with the primer adapter of sequence[5'.
GAGAGAGAGAGGATCCCAAGAGCTCAATTAATTAATAACCCCCCCCCC 3']. cDNA was
cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end:
XhoI. Host: SOLR.

Location/Qualifiers
1..1137
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="MGD:MGI:189492"
/db_xref="taxon:10090"
/clone="0610011119"
/sex="male"
/tissue_type="kidney"
/cdonc_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
99..893
/gene="Capn10"
99..893
/gene="Capn10"
/note="calpain 10
data source:MGD, source key:MGI:1344392, evidence:ISS
putative"
/codon_start=1
/protein_id="BAB22179.1"
/db_xref="GI:12832610"
/translation="NRVRAETPARELFDAAPASDSSLFVNLSTPLAQFREDITWR
RPEICATPOLFPDPNWEGQYKQGLLGCWFLCACALQKSOHLLDQVFPQCPGWS
QKYQGFCTRCRWQFHWEEVTIDRLPLCAGRLCFRCORDEVFWLPLEKAYKVHG
SYEHLWAGADALVDLTGSUAERSLKDVTKAGSQDQRPSSGSGTRCOLLHLKDR
LISCSVLSPRATAMATPTSTISLQKRKSVLSERLSWAGVLLPLP"

```

polyA_signal 1112..1117
/note="putative"
polyA_site 1137
/note="putative"
BASE COUNT 258 a 313 c 322 g 244 t
ORIGIN

alignment_scores:
Quality: 1091.00 Length: 230
Ratio: 5.005 Gaps: 0
Percent Similarity: 94.783 Percent Identity: 86.087

alignment_block:
US-09-768-877-2 x AK002548 ..

Align seg 1/1 to: AK002548 from: 1 to: 1137

```

1 MetArgLaGlyArgGlyAlaThrProAlaArgGluLeuPheArgAspAl 17
|||||
99 ATGGGGGGCGGCGCGGAGACCGCGCGGCGGAGCTCTCCGCGACGC 148
17 aAlaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrPro 34
|||||
149 GGCATTCGCCCGCCGCGACCTCTCTTTACACACTTGTCCAGCTC 198
34 euAlaGlnPheArgGluAspIleThrTrpArgArgProGlnGluIleCys 50
|||||
199 TGGCCCGATTCCGGAGACATCCTTGAGACGACCCAGAAATCTGT 248
51 AlaThrProArgLeuPheProAspAspProArgGluGlnValIysG1 67
|||||
249 GCCACACTCAGCTGTTCCAGATACACCGAGGAGGAGGAGGAGGAGCA 298
67 nGlyLeuLeuGlyAspCysTrpPheLeuCysAlaCysAlaAlaLeuGln 84
|||||
299 AGGCGCTGGGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 348
84 ySerArgHisLeuLeuAspGlnValIleProGlnGlnProSerTrp 100
|||||
349 AGATCAACACCTCTGAGACAGCTCTCCCTCAGACAGCCAGCGCTGG 398
101 AlaSpGlnGluTrpArgGlySerPheThrCysArgIleTrpGlnPheG1 117
|||||
399 TCTGACCAAAATACCAAGCTCTTCTCCTGCGGATTTGGCAGTTGG 448
117 yArgTrpValGluValThrThrAspAspArgGluProCysLeuAlaG1 134
|||||
449 ACACCTGGAGAGATGACCATGATGATGATGATGATGATGATGATG 498
134 rGluCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
|||||
499 GACTCTGCTTTCCCGGTCGACAGAGAGAGATGTTGGCGCTTCCCTA 548
151 LeuGlnValValTrpAlaValValHisGlySerTrpGlnHisLeuTrpAl 167
|||||
549 CTGGAAAGAGCCATATGCTAAGGTCATGATGATGATGATGATGATG 598
167 aGlyGlnValAlaAspAlaLeuValAspLeuPheTrpGlyLeuAlaGlu 184
|||||
599 AGGCGCAGAGGAGATGCTTACTGATGATGATGATGATGATGATGAT 648
184 rGTrpAsnLeuLysGlyValAlaGlySerGlyGlnGlnAspArgPro 200
|||||
649 GCTGGAGCTTGAAGATGACGAAAGCCAGAGGCGGAGAGAGAGAGACC 698
201 GlyArgTrpGlnHisArgTrpCysArgGlnLeuHisLeuLysAspG1 217
|||||
699 ACTGTGGGAGCAGACGAACTTGTGCGCAGCTACTCCAGTGAAGAGCG 748
217 nCysLeuIleSerCysValLeuSerProArgAlaGly 230
|||||
749 GTGTCTAAATCAGCTCTCTGTGCTTAAAGCCCGAGAGAGGAT 788

```

seq_name: gb_hlc:AK013497

seq_documentation_block:
LOCUS AK013497 1269 bp mRNA linear HTC 19-JAN-2002
DEFINITION Mus musculus adult male hippocampus cDNA, RIKEN full-length
enriched library, clone:2900008C09:calpain 10, full insert
sequence.

ACCESSION AK013497
VERSION AK013497.1 GI:12850882
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) adult male hippocampus cDNA to mRNA,
clone.lib:RIKEN full-length enriched mouse cDNA library
clone:2900008C09.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (sites)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
10349636

REFERENCE

2 (sites)
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
11042159

JOURNAL

PUBMED

REFERENCE

AUTHORS

3 (sites)
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Kono, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, I., Kita, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861

JOURNAL

PUBMED

REFERENCE

AUTHORS

4 (sites)
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5 (bases 1 to 1269)

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

Submitted (10-JUL-2000) Yoshihide Hayashizaki. The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
Kanagawa 230-0045, Japan (E-mail:genome-res@gsr.riken.go.jp,
URL: <http://genome.gsc.riken.go.jp/>, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for

g/cm^3


```

/tissue_type="pooled lung tumors"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pRT30-Pac (Pharmacia) with a
modified polylinker; Site_1: NotI; Site_2: EcoRI; 1st
strand cDNA was prepared from mRNA obtained from pooled
lung tumors with a Not I - oligo(dT) primer [5',
TGTACCAATCTGAGAGGAGCGCCCTCTTTTCTTTTCTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pRT30 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."

```

```

BASE COUNT      216 a      289 c      304 g      219 t
ORIGIN

```

```

alignment_scores:
  Quality: 1078.00      Length: 231
  Ratio: 4.945          Gaps: 1
  Percent Similarity: 94.372      Percent Identity: 85.714

```

```

alignment_block:
US-09-768-877-2 x B1410486 ..

```

```

Align seg 1/1 to: B1410486 from: 1 to: 1028

```

```

1 MetArgAlaGlyAArgGlyAlaPheProAlaArgGluLeuPheArgAspAl 17
  |||||
93 ATGCGGGCGGTCGGGCGGAGAGCGGCGGCGGAGCTTCCGGGAGCC 142
17 AlaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrProL 34
  |||||
143 GGCATTCCCGCGGCTGCTCTCTTTACACTTGTCCAGCCCTC 192
34 euaAlaGlnPheArgGluAspIleThrPArgArgProGlnIleCys 50
  |||||
193 TGGCCAGTTTCGGAGAGACATCACTGGAGACGCCAGAAATCTGT 242
51 AlaThrProArgLeuPheProAspAspProArgGluGlnValLysGln 67
  |||||
243 GCCACACTTCAGCTGTTCCAGATACCCATGCGGAGAGAGTGAACCA 292
67 nGlyLeuLeuGlyAspCysTrpPheLeuCysAlaCysAlaLeuGlnL 84
293 AGGGCTGCTGGAGATTCTGCTCTCTGTGCTGCTGCTGCTGCTGCTG 342
84 ySserAlaGlnLeuLeuAspGlnValIleProProGlnProSerTrp 100
  |||||
343 AGAGTCAACACCTCTCTGACAGGCTCTCTCCAGAGACAGCCAGCTG 392
101 AlaAspGlnGluTrpArgGlySerPheThrCysArgIleTrpGlnPhe 117
  |||||
393 TCTACCAAGAAATACAGAGCTTCTCACTGTGAGATTGGCAGTTGG 442
117 yArgTrpValGlnValThrThrAspAspArgLeuProCysLeuAlaGly 134
  |||||
443 ACACTGGAGAGAGTGAACCATAGATGATCCTCTCTGCTTCTCCGGGA 492
134 rGluLeuCysPheSerArgCysGlnArgGluAspValPheTrpLeuPro 150
493 GACTCTCTCTTCCCGGTCAGAGAGAGATGTTCTGGCTTCCCTTA 542
151 LeuGlnLysValTyrAlaLysValHisGlySerTrpGlnLysLeuTrpAl 167
  |||||
543 CTGGAAGAAAGCCCTATGCTTAAGGTCCATGATGATGAGACCTGTGG 592
167 agLysGlnValAlaAspAlaLeuValaPheLeuThrGlyLeuAlaGln 184
  |||||
593 AGGGCAAGTGCAGATGCTTAAGTATCTCCTGGAAGCTGACAGAAA 642
184 rGTrpAsnLeuLysGlyValAlaGlySerGlyGlnGlnAspArgPro 200
  |||||
643 GGTGGAGCTTGAAGATGTAACGAAAGCCAGCGGCCAGAGACAGACC 692

```

```

201 GlyArgTrpGlnHisArgTrpCysArgGlnLeuLeuHisLeuLysAsp 217
  |||||
693 AGGTCGGGAGAGCAGACACTTGTCCGACACTCTCCACTGAGAGACC 742
217 IncCysLeuLeuSerCysCysValLeuSerProArgAlaGly 230
  |||||
743 GGTGTCTATTCAGCTCTCTGCTCTTACGCCAGACAGAGCT 783
seq_name: gb_est2:BE390603

```

```

seq_documentation_block:
LOCUS      BE390603                      601 bp    mRNA    linear    EST 21-JUL-2000
DEFINITION 601284961F1 NIH_MGC_44 Homo sapiens cDNA clone IMAGE:360638 5',
            mRNA sequence.
ACCESSION  BE390603
VERSION    BE390603.1 GI:9335968
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 601)
            NIH-MGC http://mgi.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgaaps-remail.nih.gov
            Tissue Procurement: ATCC
            CDNA Library Preparation: Ling Hong/Rubin Laboratory
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LNL at:
            http://image.lnl.gov
            Plate: LNCM256 row: 3 column: 15
            High quality sequence stop: 600.
            Location/Qualifiers
                1..601
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="IMAGE:360638"
                    /clone_lib="NIH_MGC_44"
                    /tissue_type="endometrium, adenocarcinoma cell line"
                    /lab_host="DH10B (phage-resistant)"
                    /note="Organ: uterus; Vector: pRT30; Site_1: XhoI; Site_2:
                    EcoRI; cDNA made by oligo-dT priming. Directionally
                    cloned into EcoRI/XhoI sites using the following 5'
                    adaptor: GGCACGAG(C). Library constructed by Ling Hong
                    in the laboratory of Gerald M. Rubin (University of
                    California, Berkeley) using ZAP-cDNA synthesis kit
                    (Stratagene) and Superscript II RT (Life Technologies)."

```

```

BASE COUNT      104 a      173 c      225 g      99 t
ORIGIN

```

```

alignment_scores:
  Quality: 1076.00      Length: 197
  Ratio: 5.462          Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 100.000

```

```

alignment_block:
US-09-768-877-2 x BE390603 ..

```

```

Align seg 1/1 to: BE390603 from: 1 to: 601

```

```

200 ProGlyArgTrpGlnHisArgTrpCysArgGlnLeuLeuHisLeuLys 216
  |||||
3 CCAGGCCGCTGGAGAGCAGAGACTTGTCCGACAGTCTCCACCTGAAGGA 52
216 pGlnCysLeuLysSerCysValLeuSerProArgAlaGlyAlaArg 233
  |||||
53 CCAGTGTCTGATCAGCTGTGCTGCTCAGCCAGAGAGAGGTCCCGGG 102

```

233 luLeuGlyGluPheHisAlaPheIleValSerAspLeuArgGluLeuGln 249
 103 AGCTGGGGGAGTTCCATGCTTCATGTCTCGGACCTCGGGAGCTCCAG 152
 250 GlyAlaGlyGlnCysIleLeuLeuArgIleGlnAsnProIrrpGI 266
 153 GGTACGGGGGGCAGTGCATCTCTGCTGCGGATCCAGAACCCCTGGGG 202
 266 yArgArgCysTrpGlnGlyLeuTrpArgGluGlyGlyGluGlyTrpSerG 283
 203 CCGCGGTGCTGGCAGGGGCTCTGGAGAGGGGGGTGAAGGTGGAGCC 252
 283 InValAspAlaValAlaSerGluLeuLeuSerGlnLeuGlnGluGly 299
 253 AGGTAGATGCACGGGTAGCATCTGAGTCTCTGCCAGCTCCAGGAAGG 302
 300 GluPheTrpValGluGluGluGluPheLeuArgGluPheAspGluLeuTh 316
 303 GAGTTCTGGTGGAGGAGGAGTCTCTCAGGGAGTTGACGAGGCTCAC 352
 316 rValGlyTrpProValThrGluAlaGlyHisLeuGlnSerLeuTrpThrG 333
 353 CGTTGGCTACCGGTCACGGAGGGCGGCCACCTGCAGAGGCTCTACACAG 402
 333 luArgLeuLeuCysHisThrArgAlaLeuProGlyAlaTrpValLysGly 349
 403 AGAGGCTGCTCTGCCATACGGGGCGCTGCCTGGGGCTGGGTCAAGGCG 452
 350 GlnSerAlaGlyGlyCysArgAsnAsnSerGlyPheProSerAsnProLy 366
 453 CAGTCAGCAGGAGGCTGCCGGAACAACAGCGGGTCTCCCAAGCAACCCAA 502
 366 sPheTrpLeuArgValSerGluProSerGluValTyrIleAlaValLeuG 383
 503 ATTCTGGTGGGGTCTCAGAACCGAGTGAGGTGTACATTGGCTCTCTGC 552
 383 InArgSerArgLeuHisAlaAlaAspTrpAlaGlyArgAla 396
 553 AGAGATCCAGGCTGCACGCGGAGCTGGGCAAGCGCGGGCC 593

name: qb_est2:BI412050

seq_documentation_block:	
LOCUS	BI412050 940 bp mRNA linear EST 14-AUG-2001
DEFINITION	G02965960F1 NCI_CGAP_Lu33 Mus musculus cdna clone IMAGE:5121446 5', mRNA sequence.
ACCESSION	BI412050
VERSION	BI412050.1 GI:15172973
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE	1 (bases 1 to 940)
JOURNAL	NIH-MGSC http://mgc.nci.nih.gov/ .
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)
Contact:	Robert Strausberg, Ph.D.
Email:	cgaps-remail.nih.gov
Tissue Procurement:	Gilbert Smith, Ph.D.
CDNA Library Preparation:	M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
CDNA Library Arrayed by:	The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by:	Incyte Genomics, Inc.
Clone distribution:	NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate:	LLAM1296 row: g column: 15
High quality sequence start:	3
High quality sequence stop:	858.
Location/Qualifiers	1..940
FEATURES	source

```

/organism="Mus musculus"
/strain="Czech II"
/db_xref="taxon:10090"
/clone_image="IMAGE:5121446"
/clone_lib="NCI_CGAP_Lu33"
/tissue_type="pooled lung tumors"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: p7T3D-Pac (Pharmacia) with a
modified polylinker; Site-1: NotI; Site-2: EcoRI; 1st
strand cDNA was prepared from mRNA obtained from pooled
lung tumors with a NotI - oligo(dT) primer [5',
TGTATCAACTGTAAGTGGAGCGCGCCCTCTTTTTTTTTTTT 3']."
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with NotI and cloned into the NotI
I and Eco RI sites of the modified p7T3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo.
193 a 269 c 275 g 203 t

```

**BASE COUNT
ORIGIN**

alignment_scores:

Quality:	1068.50	Length:	289
Ratio:	4.586	Gaps:	4
Percent Similarity:	80.623	Percent Identity:	70.934

alignment_block:

US-09-768-877-2 x BI412050

Align seg 1/1 to: BI412050 from: 1 to: 940

1 MetArgAlaGlyVArgGlyValAlaThrProAlaArgGluLeuPheArgAspAl 17
|||||
58 ATGCGGGCGGTCCGGGGCGGAGACCCCGCGCGGAGGCTCTTCGGGACGC 107
17 aaAlaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrProL 34
|||||
108 GGCATTCCCGGCTCGGACTCCTCGCTCTTTTACAACACTTGTCCACGCCCTC 157
34 euAlaGlnPheArgGluAspIleThrTrpArgProGlnGluIleCys 50
|||||
158 TGGCCCGAGTTTCGGGAGGACATCACTTGGAGACAGCCCAAGSAAATCTGT 207
51 AlaThrProArgLeuPheProAspAspProArgGluGlyGlnValLysG 67
|||||
208 GCCACACCTCAGCTGTTTCCAGATAACCCATGTGGGAGGACAGGTGAAGCA 257
67 nGlyLeuLeuGlyAspCysTrpPheLeuCysAlaCysAlaAlaLeuGlnL 84
|||||
258 AGGGCTGCTGGGAGATTGCTGGTCCCTGTGGCTGTGGCGGCCCTTCAGA 307
84 ysSerArgHisLeuLeuAspGlnValIleProProGlyGlnProSerTrp 100
|||||
308 AGAGTCAACACCTCCTGGACAGGCTCTCCCTCCAGGACACCGCGCTGG 357
101 AlaAspGlnGluTyrArgGlySerPheThrCysArgIleTrpGlnPheG 117
|||||
358 TCTGACCAAGAAATACCAAGGCTCTTCACCTGCGGATTTGGCAGTTTGG 407
117 YArgTrpValGluValThrThrAspAspArgLeuProCysLeuAlaGlyA 134
|||||
408 ACACTGGGAGGAGTGACCATAGATGATGCTGTGCCTGTCTTGCCCGGA 457
134 rGluCysPheSerArgCysGlnArgGluAspValPheTrpLeuProLeu 150
|||||
458 GACTCTGCTTTTCCCGGTGCCAGAGAGAGAGATGTGTGTGGCTTCCTCTTA 507
151 LeuGluLysValTyrAlaLysValHisGlySerTyrGluHisLeuTrpAl 167
|||||
508 CTGGAAAAGGCCATGTCTAAGTCCATGGATCGTATGACGACCTGTGGGC 557
167 aGlyGlnValAlaAlaAspAlaLeuValAspLeuThrGlyGlyLeuAlaGluA 184
|||||

```

558 AGGCAAGTGGCAGATGCTTAGTGAGATCTCACTGCAAGCTGCGAGAAA 607
184 TGTPrAsnLeuLysGly.ValAlaGlySerGlyGlyGlnGlnAspArgPr 200
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
608 GGTGGACGCTTGACAGATGTACGAAAGCAGCCGCGCAGACAGACAGACC 657
200 OGlyArgTPrGlnHisArgThrCysArgGlnLeuLeuHisLeuLysAsp 217
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
658 CAGTGTGGGAGACAGAACTGTGCGCAGCTACTCCACTGGAAGAGACC 707
217 LncCysLeuLeuSerCysCysValLeuSer.ProArgAlaGly..... 230
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
708 GGTGTCTAATCAGCTGCTGTGTGTAGCCCGCCAGAGCAGGTACAGCTAT 757
231 .....AlaArgGlnLeuGly.. 235
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
758 GGCTACACTCCCTCCACACATTAGCTGTGTACAGACGAAAGAAATCGGTC 807
235 ..... 235
808 CTTGAGAAAGCTGTCTGTGGTGGGTGTCTGTGTACCCCTCTAG 857
236 .....GlnPheHisAlaPheIleVal 242
858 AGAGAACTATACACTTCTCTCAGACTTATCTTCTTACAGCTAGGCGCA 907
242 LserAspLeuArg 246
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
908 AGCGACCTACGT 920
seq_name: gb_est2:BE619302

seq_documentation_block:
LOCUS BE619302 789 bp mRNA linear EST 20-OCT-2000
DEFINITION 601473184F1 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:3876141 5',
mRNA sequence.
ACCESSION BE619302
VERSION BE619302.1 GI:9890240
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 789)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/DTF/Gazdar
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LAM9636 row: g column: 22
High quality sequence stop: 663.
Location/Qualifiers
1..789
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3876141"
/clone_lib="NIH_MGC_68"
/tissue_type="large cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."
BASE COUNT 145 a 218 c 295 g 131 t
ORIGIN

```

```

alignment_scores:
    Quality: 1056.50      Length: 251
    Ratio: 4.439         Gaps: 7
    Percent Similarity: 94.821    Percent Identity: 92.430

alignment_block:
US-09-768-877-2 x BE619302
Align seg 1/1 to: BE619302 from: 1 to: 789

36 GlnPheArgGlnAspIleThrTPrArgPrProGlnGlnIleCysAlaLeu 52
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
13 CAGTCCCGCAGAGACATCAGCTGGAGCGCGCCCGCAGAGAT.TGTGCCAC 61
52 TPrArgLeuPheProAspAspProArgGlnGlnValLysGlnGlyL 69
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
62 ACCCGGCTGTTCAGATGACCCAGCGGAGAGGCGAGGTGAAGCAGGGCC 111
69 eulGlyAspCysTPrPheLeuLysAlaCysAlaAlaLeuGlnLysSer 85
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
112 TCTGGGGGA.TGCTGCTTCTGTGTGCTGCTGCGCGCTCAGAAAGAC 160
86 ArgHisLeuLeuAspGlnValIleProProGlnGlnProSerTPrAla 102
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
161 AGGACCTCTCGACACAGATCCTCTCCGGACAGCCGAGCTGGCGCA 210
102 pGlnGlnTPrArgLysSerPheThrCysArgIleTPrPheGlnArgT 119
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
211 CCAGGAGTACCGGGGCTCTTCACTGTCGATTTGGACAGT.GGAGCGT 259
119 TPrValGlnValThrThrAspAspArgLeuProCysLeuAlaGlyArg 135
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
260 GGGTGGAGGTACACAGATATCCGCTGCGCTTGCCTTGCAGGAGACTC 309
136 CysPheSerArgCysGlnArgGlnAspValPheThrLeuProLeuLeu 152
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
310 TGTTCCTCCGCTGCGACAGGAGAGATGTCTTCCCTCCCTTACT.GA 358
152 ulysValTyrAlaLysValHisGlySerTyrGlnHisLeuTPrAla 169
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
359 AAAGCTTACCCCAAGCTCCATGGTCTTACGACACCTGTGGGCGGCG 408
169 lncValAlaAspAlaLeuValAspLeuThrGlyLeuAlaGlnArgTPr 185
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
409 AGGTGGGAGATGCCCTGTGACTGACCGCGGCTG.GCAGAAAGATGG 457
186 AsnLeuLysGlyValAlaGlySerGlyGlnGlnAspArgPrProGly 202
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
458 AACCTGAAGGCGCTTACAGAGAGCGGAGCGCAGAGACAGCGCGCG 507
202 gTPrGlnHisArgThrCysArgGlnLeuLeuHisLeuLys.AspGlnCys 218
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
508 CTGGAG.CACAGACTTGTGCGCAGCTGCTCCACTTAAGGACCAAGGT 556
219 LeuLeuSerCysCysValLeuSerProArg.AlaGlyAlaArgGlnLeu 235
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
557 CTGATCAGCTGCTGCGCTGCTGACCCCAAGAGAGGTGCCCGGAGCTGG 606
235 LysGlnPheHisAlaPheIleValSerAspLeuArgGlnLeuGlnGly 251
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
607 GGAAGTTCCTATCTTATTTCTCGAGCTCGGGAAGCTCCAGGGTCA 656
251 nAlaGlyGlnCysIleLeuLeuLeu.ArgIleGlnAsnProTPrGlyArg 267
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
657 GGGGGGCACTGCATC...CTGCTGTGGATCAGAAACCG...GGGCG 700
268 ArgCysTPrGlnGlnLeuTPrArgGlnGlyGlnGlyLysPsr 282
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
701 GGGGTGTGCGCAGGGGCTG...GAAAGGGGGTGAAGGGTGGGCG 742
seq_name: gb_est1:AL558905

```

```

seq_documentation_block:
LOCUS       AL558905 LTI_NFL008_TC2 Homo sapiens cdna clone CS0DJ007YG13.5
DEFINITION  prime, mRNA sequence.
ACCESSION   AL558905
VERSION     AL558905.1 GI:12903884
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 776)
  Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  Full-length cDNA libraries and normalization
  Unpublished (2001)
  Contact: Genoscope
  Genoscope - Centre National de Sequencage
  BP 191 91006 EVRY cedex - France
  Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
  Location/Qualifiers
    1..776
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="CS0DJ007YG13"
      /clone_lib="LTI_NFL008_TC2"
      /sex="male"
      /tissue_type="T cells from T cell leukemia"
      /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : flliang@lifetech.com URL :
http://fulllength.invitrogen.com"
119 a 237 c 292 g 124 t 4 others
BASE COUNT
ORIGIN

```

```

alignment_scores:      Length: 196
                        Quality: 1047.00
                        Ratio: 5.425
                        Gaps: 0
                        Percent Similarity: 98.469
                        Percent Identity: 97.959

alignment_block:
  US-09-768-877-2 x AL558905

  Align seg 1/1 to: AL558905 from: 1 to: 776

1 MetArgAlaGlyArgGjYAlaThrProAlaArgGluLeuPheArgAspAl 17
187 ATCGGGCGGGCGGGGGCGGAGCCGCCGGAGGGAGCTGTTCGGGACGC 236
17 aAlaPheProAlaAlaAspSerSerLeuPheCysAspLeuSerThrProL 34
237 CGCCTTCCCGCGCGGACTCCTCGCTCTTCTGGGACTTGTCTACGCCGC 286
34 euAlaGlnPheArgGluAspIleThrTrpArgArgProGlnGluIleCys 50
287 TGGGCCACAGTTCCGGCAGGAGATCACGTGGAGGGGGCCCCAGGAGTTGT 336
51 AlaThrProArgLeuPheProAspAspProArgGluClyGlnVallysG1 67
337 GCCACACCCCGGTGTTTCCAGATGACCCACGGGAGGGCAGGTGAAGCA 386
67 nGlyLeuLeuGlyAspCysTrpPheLeuCysAlaCysAlaAlaLeuGlnL 84
387 GGGCTGCTGGGGGATTGCTGGTTCCTCTGTGTGCTTCGCCACCGCTGCAGA 436
84 ysSerArgHisLeuLeuAspGlnValIleProProGlyGlnProSerTrp 100

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2002, 10:34:46 ; Search time 14.65 Seconds
(without alignments)
1120.409 Million cell updates/sec

Title: US-09-768-877-2

Perfect score: 3649
Sequence: 1 MRAGCATPARELFRDAAP.....SOENIGOLFQEVSVAAVMT 672

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/prodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/prodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/prodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/prodata/2/1aa/ACTUS.COMB.pep:*
6: /cgn2_6/prodata/2/1aa/Dackfilest.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3649	100.0	672	US-09-422-869-2	Sequence 2, Appl1
2	2963	81.2	666	US-09-422-869-18	Sequence 18, Appl1
3	2715.5	74.4	517	US-09-422-869-6	Sequence 6, Appl1
4	2701.5	74.0	544	US-09-422-869-4	Sequence 4, Appl1
5	2699	74.0	513	US-09-422-869-8	Sequence 8, Appl1
6	2336	64.0	444	US-09-422-869-10	Sequence 10, Appl1
7	858.5	23.5	274	US-09-422-869-12	Sequence 12, Appl1
8	635	17.4	703	US-08-835-099A-1	Sequence 1, Appl1
9	635	17.4	703	US-09-157-349-1	Sequence 2, Appl1
10	635	17.4	712	US-08-835-099A-2	Sequence 1, Appl1
11	635	17.4	712	US-09-157-349-2	Sequence 2, Appl1
12	618.5	16.9	821	US-09-422-869-22	Sequence 22, Appl1
13	618	16.9	821	US-09-422-869-24	Sequence 24, Appl1
14	612	16.8	703	US-09-422-869-27	Sequence 27, Appl1
15	604	16.6	700	US-08-726-525-7	Sequence 7, Appl1
16	604	16.6	700	US-08-487-942-7	Sequence 7, Appl1
17	604	16.6	700	US-08-726-036A-7	Sequence 7, Appl1
18	604	16.6	700	US-08-423-869-23	Sequence 23, Appl1
19	604	16.6	700	US-09-083-516-7	Sequence 7, Appl1
20	584	16.0	639	US-09-422-869-25	Sequence 25, Appl1
21	561.5	15.4	690	US-09-422-869-28	Sequence 28, Appl1
22	513.5	14.1	139	US-09-422-869-14	Sequence 14, Appl1
23	483.5	13.3	641	US-09-422-869-26	Sequence 26, Appl1
24	472	12.9	138	US-09-422-869-16	Sequence 16, Appl1
25	271	7.4	854	US-08-928-692-17	Sequence 17, Appl1
26	264.5	7.2	842	US-08-928-692-18	Sequence 18, Appl1
27	133	3.6	67	US-09-120-365-97	Sequence 97, Appl1

28	133	3.6	67	4	US-09-515-039-97	Sequence 97, Appl1
29	124	3.4	67	3	US-09-120-365-95	Sequence 95, Appl1
30	124	3.4	67	4	US-09-515-039-95	Sequence 95, Appl1
31	121	3.3	67	3	US-09-120-365-98	Sequence 98, Appl1
32	121	3.3	67	4	US-09-515-039-98	Sequence 98, Appl1
33	117.5	3.2	67	3	US-09-120-365-99	Sequence 99, Appl1
34	117.5	3.2	67	4	US-09-515-039-99	Sequence 99, Appl1
35	117	3.2	67	3	US-09-120-365-96	Sequence 96, Appl1
36	117	3.2	67	4	US-09-515-039-96	Sequence 96, Appl1
37	115	3.2	67	3	US-09-120-365-94	Sequence 94, Appl1
38	115	3.2	67	4	US-09-515-039-94	Sequence 94, Appl1
39	110	3.0	3519	4	US-08-804-227C-9	Sequence 4, Appl1
40	105	2.9	1996	2	US-08-804-198-3	Sequence 9, Appl1
41	105	2.9	1996	2	US-08-804-198-3	Sequence 9, Appl1
42	102.5	2.8	512	4	US-09-356-818A-2	Sequence 2, Appl1
43	102.5	2.8	548	1	US-08-247-902A-2	Sequence 2, Appl1
44	102.5	2.8	687	5	PCT-US91-09784-2	Sequence 2, Appl1
45	100.5	2.8	505	1	US-08-220-603A-10	Sequence 10, Appl1

ALIGNMENTS

RESULT 1
US-09-422-869-2
Sequence 2, Application US/09422869
Patent No. 6235481
GENERAL INFORMATION:
APPLICANT: POLONSKY, KENNETH S.
APPLICANT: HORIKAWA, YUKIO
APPLICANT: COX, NANCY J.
APPLICANT: SRENNAN, SEAMUS
APPLICANT: ZHOU, YUN-PING
APPLICANT: OTANI, KENICHI
APPLICANT: HANIS, CRAIG L.
APPLICANT: BELL, GRAEME I.
TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
FILE REFERENCE: AKD:307
CURRENT APPLICATION NUMBER: US/09/422,869
CURRENT FILING DATE: 1999-10-21
EARLIER APPLICATION NUMBER: 60/134,175
EARLIER FILING DATE: 1999-05-13
NUMBER OF SEQ ID NOS: 30
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 2
LENGTH: 672
TYPE: PRT
ORGANISM: Human
US-09-422-869-2

Query Match 100.0%: Score 3649; DB 4; Length 672;
Best Local Similarity 100.0%: Pred. No. 0;
Matches 672; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MRAGCATPARELFRDAAPADSSIFCDLSTPLAOFREDITMRROEICATPRLEPPDP 60
1 MRAGCATPARELFRDAAPADSSIFCDLSTPLAOFREDITMRROEICATPRLEPPDP 60
61 REGGVKQGLGSCWELCAALQKSHLLDQYIPGQPSMAQOEXRGSEFCRIMQGRMV 120
61 REGGVKQGLGSCWELCAALQKSHLLDQYIPGQPSMAQOEXRGSEFCRIMQGRMV 120
121 EYTDORLPCFLAGRCFSRCOREDFWMLPLEKKYAVKGVSTEHLMAGQVADALVDLGG 180
121 EYTDORLPCFLAGRCFSRCOREDFWMLPLEKKYAVKGVSTEHLMAGQVADALVDLGG 180
181 IAEARNLKGVAAGSGGQDPGRMEHRTCRQLHLMDQCLISCCVSPRAGARELGEFFNAF 240
181 IAEARNLKGVAAGSGGQDPGRMEHRTCRQLHLMDQCLISCCVSPRAGARELGEFFNAF 240
241 IYSDRELQGAQGCITLLIRIONPWGRRCWGLMBGEGMSQVDAVAASELLSOLQEGE 300

Db 241 IVSDRELQAGQACILLRIQNPGRRCWQGLWREGGEGNSQVDAVASELLSQLEGE 300
QY 301 FWVEEELREFDELTVGYPVTEAGHLQSLYTERLLCHTRALPGAWKQSGAGGCRNNSG 360
Db 301 FWVEEELREFDELTVGYPVTEAGHLQSLYTERLLCHTRALPGAWKQSGAGGCRNNSG 360
QY 361 FPSNPKFWLRVSEPEVYIAVLQSRRLHAADWAGARALVGDSTSWSPASIPGKHQYAV 420
Db 361 FPSNPKFWLRVSEPEVYIAVLQSRRLHAADWAGARALVGDSTSWSPASIPGKHQYAV 420
QY 421 GLHLKVKERRVNLPRVLSMPVAGTACHAYDREHVLRCLESPGYLLAVPSTFLKDAPGE 480
Db 421 GLHLKVKERRVNLPRVLSMPVAGTACHAYDREHVLRCLESPGYLLAVPSTFLKDAPGE 480
QY 481 FLLRVFSTGRVLSAIRAVAKNTTPCAALPAGEWGTQVLRGSRWYQOTAGGSRNFASYPT 540
Db 481 FLLRVFSTGRVLSAIRAVAKNTTPCAALPAGEWGTQVLRGSRWYQOTAGGSRNFASYPT 540
QY 541 NCFPFSPVEGPGPCVRITLHOCRPSDTEPHPIGFHIFQVPEGGRSQDAPPLLOEPL 600
Db 541 NCFPFSPVEGPGPCVRITLHOCRPSDTEPHPIGFHIFQVPEGGRSQDAPPLLOEPL 600
QY 601 LSCVPHRYAQEVSRLLCPAGTYKVPSTYLPDTEGAFVTIATRIDRPSIHSQEMLGOF 660
Db 601 LSCVPHRYAQEVSRLLCPAGTYKVPSTYLPDTEGAFVTIATRIDRPSIHSQEMLGOF 660
QY 661 LOEVSVMVMK 672
Db 661 LOEVSVMVMK 672

RESULT 2
US-09-422-869-18
; Sequence 18, Application US/09422869
; Patent No. 6235481
; GENERAL INFORMATION:
; APPLICANT: POLONSKY, KENNETH S.
; APPLICANT: HORIKAWA, YUKIO
; APPLICANT: ODA, NAOHISA
; APPLICANT: COX, NANCY J.
; APPLICANT: SREENAN, SEAMUS
; APPLICANT: ZHOU, YUN-PING
; APPLICANT: OTANI, KENICHI
; APPLICANT: HANIS, CRAIG L.
; APPLICANT: BELL, GRAEME I.
; TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
; FILE REFERENCE: ARCD:307
; CURRENT APPLICATION NUMBER: US/09/422,869
; EARLIER FILING DATE: 1999-10-21
; EARLIER FILING DATE: 1999-05-13
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 666
; TYPE: PRT
; ORGANISM: Mus musculus.
US-09-422-869-18

Query Match 81.2%; Score 2963; DB 4; Length 666;
Best Local Similarity 81.1%; Pred. No. 3.7e-289;
Matches 544; Conservative 45; Mismatches 76; Indels 6; Gaps 2;

QY 1 MRAGATPARELFDAAPADSSLFCDLSTPLAQFREDITWRPQICATPRLFPDPP 60
Db 1 MRVRAETPARELFDAAPADSSLFYNLSTPLAQFREDITWRPQICATPQLFPDPP 60
QY 61 REGQVKGGLGDCWFLCACLQKSRHLLDQVIPPQGSWADQYRGSTCRIMQGRWY 120
Db 61 WEGQVKGGLGDCWFLCACLQKSRHLLDQVIPPQGSWADQYRGSTCRIMQGRWY 120

QY 121 EVTTDDRLPCLAGRLCFRCQREDVFWLPLEKVKYAKVHGSYEHLMWAGQVADALVDLTGG 180
Db 121 EVTTDDRLPCLAGRLCFRCQREDVFWLPLEKVKYAKVHGSYEHLMWAGQVADALVDLTGG 180
QY 181 LAERNLKVAGSGQGDPRGWEHRTCRQLHLKQDCLISCCVLSPRAGARELGEFHF 240
Db 181 LAERNLKVAGSGQGDPRGWEHRTCRQLHLKQDCLISCCVLSPRAGARELGEFHF 240
QY 241 IVSDRELQAGQACILLRIQNPGRRCWQGLWREGGEGNSQVDAVASELLSQLEGE 300
Db 241 IVSDRELQAGQACILLRIQNPGRRCWQGLWREGGEGNSQVDAVASELLSQLEGE 300
QY 301 FWVEEELREFDELTVGYPVTEAGHLQSLYTERLLCHTRALPGAWKQSGAGGCRNNSG 360
Db 301 FWVEEELREFDELTVGYPVTEAGHLQSLYTERLLCHTRALPGAWKQSGAGGCRNNSG 360
QY 361 FPSNPKFWLRVSEPEVYIAVLQSRRLHAADWAGARALVGDSTSWSPASIPGKHQYAV 420
Db 361 FPSNPKFWLRVSEPEVYIAVLQSRRLHAADWAGARALVGDSTSWSPASIPGKHQYAV 420
QY 421 GLHLKVKERRVNLPRVLSMPVAGTACHAYDREHVLRCLESPGYLLAVPSTFLKDAPGE 480
Db 421 GLHLKVKERRVNLPRVLSMPVAGTACHAYDREHVLRCLESPGYLLAVPSTFLKDAPGE 480
QY 481 FLLRVFSTGRVLSAIRAVAKNTTPCAALPAGEWGTQVLRGSRWYQOTAGGSRNFASYPT 540
Db 481 FLLRVFSTGRVLSAIRAVAKNTTPCAALPAGEWGTQVLRGSRWYQOTAGGSRNFASYPT 540
QY 541 NCFPFSPVEGPGPCVRITLHOCRPSDTEPHPIGFHIFQVPEGGRSQDAPPLLOEPL 600
Db 541 NCFPFSPVEGPGPCVRITLHOCRPSDTEPHPIGFHIFQVPEGGRSQDAPPLLOEPL 600
QY 601 LSCVPHRYAQEVSRLLCPAGTYKVPSTYLPDTEGAFVTIATRIDRPSIHSQEMLGOF 660
Db 601 LSCVPHRYAQEVSRLLCPAGTYKVPSTYLPDTEGAFVTIATRIDRPSIHSQEMLGOF 660
QY 661 LOEVSVMVMK 671
Db 661 LOEVSVMVMK 665

RESULT 3
US-09-422-869-6
; Sequence 6, Application US/09422869
; Patent No. 6235481
; GENERAL INFORMATION:
; APPLICANT: POLONSKY, KENNETH S.
; APPLICANT: HORIKAWA, YUKIO
; APPLICANT: ODA, NAOHISA
; APPLICANT: COX, NANCY J.
; APPLICANT: SREENAN, SEAMUS
; APPLICANT: ZHOU, YUN-PING
; APPLICANT: OTANI, KENICHI
; APPLICANT: HANIS, CRAIG L.
; APPLICANT: BELL, GRAEME I.
; TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
; FILE REFERENCE: ARCD:307
; CURRENT APPLICATION NUMBER: US/09/422,869
; EARLIER FILING DATE: 1999-10-21
; EARLIER FILING DATE: 1999-05-13
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 517
; TYPE: PRT
; ORGANISM: Human
US-09-422-869-6

Query Match 74.4%; Score 2715.5; DB 4; Length 517;
Best Local Similarity 76.8%; Pred. No. 2e-264;
Matches 516; Conservative 1; Mismatches 0; Indels 155; Gaps 1;

```

QY 1 MRAGCATPARELFRDAAFPAAADSSLCFCDLSTPLAQFREDITWRRPOICATPRLFPDDP 60
Db 1 MRAGCATPARELFRDAAFPAAADSSLCFCDLSTPLAQFREDITWRRPOICATPRLFPDDP 60
QY 61 REGOVKOGLLGDCWFLCACAALQKSRHLLDVIPPGOPSWADQYRGSGFTCRIMQFGRMV 120
Db 61 REGOVKOGLLGDCWFLCACAALQKSRHLLDVIPPGOPSWADQYRGSGFTCRIMQFGRMV 120
QY 121 EVTTDDRLPCLAGRLCFSCRCQREDVFWLPLEKRYAARVHSGYEHLMAQVADALVDTLGG 180
Db 121 EVTTDDRLPCLAGRLCFSCRCQREDVFWLPLEKRYAARVHSGYEHLMAQVADALVDTLGG 180
QY 121 EVTTDDRLPCLAGRLCFSCRCQREDVFWLPLEKRYAARVHSGYEHLMAQVADALVDTLGG 180
Db 121 EVTTDDRLPCLAGRLCFSCRCQREDVFWLPLEKRYAARVHSGYEHLMAQVADALVDTLGG 180
QY 181 LAERWNLKGVASGGQODRPGRMWHRCTROLHLKDCCLISCCVLSFRAGARELGEFHA 240
Db 181 LAERWNLKGVASGGQODRPGRMWHRCTROLHLKDCCLISCCVLSFRAGARELGEFHA 240
QY 241 IVSDRLRELQAGOCILLRIQNPWGRRCWGLMRBEGEGMSQYDAVASSELLSQLQEGE 300
Db 241 IVSDRLRELQAGOCILLRIQNPWGRRCWGLMRBEGEGMSQYDAVASSELLSQLQEGE 300
QY 301 FWVEEERFLREFDELFTGYPTAGHLOSILYTERLCHTRALPGAWKGGAGGCRNNG 360
Db 301 FWVEEERFLREFDELFTGYPTAGHLOSILYTERLCHTRALPGAWKGGAGGCRNNG 360
QY 361 FPSNPKFWLVSSESEYIYAVLQSRSLHAADWAGARALVGDSTMSPASIPGKHQAV 420
Db 361 FPSNPKFWLVSSESEYIYAVLQSRSLHAADWAGARALVGDSTMSPASIPGKHQAV 420
QY 421 GLHLMKVEKRRVNLPRVLSMPVAGTACHAYDREVHLKCLSPGYIYAVSTFLKDAPE 480
Db 421 GLHLMKVEKRRVNLPRVLSMPVAGTACHAYDREVHLKCLSPGYIYAVSTFLKDAPE 480
QY 481 FILRVSTGRVSLALRAVAKNTTGAALPAGENGTVOLGSRWVGQTAGSRNFASTPT 540
Db 481 FILRVSTGRVSLALRAVAKNTTGAALPAGENGTVOLGSRWVGQTAGSRNFASTPT 540
QY 541 NPCEPSPVPEPGPRCVRTILHCHCRPSTDEHPHIGFHIQVPEEGRSODAPRLLOEPL 600
Db 541 NPCEPSPVPEPGPRCVRTILHCHCRPSTDEHPHIGFHIQVPEEGRSODAPRLLOEPL 600
QY 601 LSCVPRYAOEVSRLCLLPAGTYKYVPSTYLPDTBGAFTYIATRIDRPSIHQEMLQGP 660
Db 601 LSCVPRYAOEVSRLCLLPAGTYKYVPSTYLPDTBGAFTYIATRIDRPSIHQEMLQGP 660
QY 661 LQEVPRYAOEVSRLCLLPAGTYKYVPSTYLPDTBGAFTYIATRIDRPSIHQEMLQGP 660
Db 661 LQEVPRYAOEVSRLCLLPAGTYKYVPSTYLPDTBGAFTYIATRIDRPSIHQEMLQGP 660

```

```

; LENGTH: 544
; TYPE: PRT
; ORGANISM: Human
US-09-422-869-4

Query Match
Best Local Similarity 74.0%; Score 2701.5; DB 4; Length 544;
Matches 498; Conservative 5; Mismatches 6; Indels 13; Gaps 1;

QY 1 MRAGCATPARELFRDAAFPAAADSSLCFCDLSTPLAQFREDITWRRPOICATPRLFPDDP 60
Db 1 MRAGCATPARELFRDAAFPAAADSSLCFCDLSTPLAQFREDITWRRPOICATPRLFPDDP 60
QY 61 REGOVKOGLLGDCWFLCACAALQKSRHLLDVIPPGOPSWADQYRGSGFTCRIMQFGRMV 120
Db 61 REGOVKOGLLGDCWFLCACAALQKSRHLLDVIPPGOPSWADQYRGSGFTCRIMQFGRMV 120
QY 121 EVTTDDRLPCLAGRLCFSCRCQREDVFWLPLEKRYAARVHSGYEHLMAQVADALVDTLGG 180
Db 121 EVTTDDRLPCLAGRLCFSCRCQREDVFWLPLEKRYAARVHSGYEHLMAQVADALVDTLGG 180
QY 121 EVTTDDRLPCLAGRLCFSCRCQREDVFWLPLEKRYAARVHSGYEHLMAQVADALVDTLGG 180
Db 121 EVTTDDRLPCLAGRLCFSCRCQREDVFWLPLEKRYAARVHSGYEHLMAQVADALVDTLGG 180
QY 181 LAERWNLKGVASGGQODRPGRMWHRCTROLHLKDCCLISCCVLSFRAGARELGEFHA 240
Db 181 LAERWNLKGVASGGQODRPGRMWHRCTROLHLKDCCLISCCVLSFRAGARELGEFHA 240
QY 241 IVSDRLRELQAGOCILLRIQNPWGRRCWGLMRBEGEGMSQYDAVASSELLSQLQEGE 300
Db 241 IVSDRLRELQAGOCILLRIQNPWGRRCWGLMRBEGEGMSQYDAVASSELLSQLQEGE 300
QY 301 FWVEEERFLREFDELFTGYPTAGHLOSILYTERLCHTRALPGAWKGGAGGCRNNG 360
Db 301 FWVEEERFLREFDELFTGYPTAGHLOSILYTERLCHTRALPGAWKGGAGGCRNNG 360
QY 361 FPSNPKFWLVSSESEYIYAVLQSRSLHAADWAGARALVGDSTMSPASIPGKHQAV 420
Db 361 FPSNPKFWLVSSESEYIYAVLQSRSLHAADWAGARALVGDSTMSPASIPGKHQAV 420
QY 421 GLHLMKVEKRRVNLPRVLSMPVAGTACHAYDREVHLKCLSPGYIYAVSTFLKDAPE 480
Db 421 GLHLMKVEKRRVNLPRVLSMPVAGTACHAYDREVHLKCLSPGYIYAVSTFLKDAPE 480
QY 481 FILRVSTGRVSLALRAVAKNTTGAALPAGENGTVOLGSRWVGQTAGSRNFASTPT 540
Db 481 FILRVSTGRVSLALRAVAKNTTGAALPAGENGTVOLGSRWVGQTAGSRNFASTPT 540

```

```

RESULT 5
US-09-422-869-8
; Sequence 8, Application US/09422869
; Patent No. 6235481
; GENERAL INFORMATION:
; APPLICANT: POLONSKY, KENNETH S.
; APPLICANT: HORIKAWA, YUKIO
; APPLICANT: ODA, NAOHISA
; APPLICANT: COX, NANCY J.
; APPLICANT: SREENAN, SEAMUS
; APPLICANT: ZHOU, YUN-PING
; APPLICANT: HANIS, CRAIG I.
; TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
; FILE REFERENCE: ARCD:307
; CURRENT FILING DATE: 1999-10-21
; EARLIER APPLICATION NUMBER: 60/134,175
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 8
; TYPE: PRT
; ORGANISM: Human

```

US-09-422-869-8

Query Match 74.0%; Score 2699; DB 4; Length 513;
Best Local Similarity 98.2%; Pred. No. 8.9e-263;
Matches 496; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 1 MRAGGATPARELFDAAFPAAADSSLCFCDLSTPLAQFREDITWRRPOEICATPRLFPDDP 60
DB 1 MRAGGATPARELFDAAFPAAADSSLCFCDLSTPLAQFREDITWRRPOEICATPRLFPDDP 60

QY 61 REGOVKGLGDCWFLCACAALQKSRHLQDVIPPGQPSWADOEYRGSTCRINQFGRW 120
DB 61 REGOVKGLGDCWFLCACAALQKSRHLQDVIPPGQPSWADOEYRGSTCRINQFGRW 120

QY 121 EVTTDDRLPCLAGRLCFRCQREDVFWLPLEKYYAKVHGSYEHLMWAGOVADALVDLTGG 180
DB 121 EVTTDDRLPCLAGRLCFRCQREDVFWLPLEKYYAKVHGSYEHLMWAGOVADALVDLTGG 180

QY 181 LAERNLKGAVSGGQODRPGWEHRTCRQLHLKQDCLISCCVLSPRAGARELGEHAF 240
DB 181 LAERNLKGAVSGGQODRPGWEHRTCRQLHLKQDCLISCCVLSPRAGARELGEHAF 240

QY 241 IVSDRLQOGAGOCILLRIONPWRCWQGLWREGGEGSWQDAVAASELLSLOQEGE 300
DB 241 IVSDRLQOGAGOCILLRIONPWRCWQGLWREGGEGSWQDAVAASELLSLOQEGE 300

QY 301 FWEEEFRLREFDELTVGYPVTEAGHLQSLYTERLLCHTRALPGAWKQSGAGCRNNSG 360
DB 301 FWEEEFRLREFDELTVGYPVTEAGHLQSLYTERLLCHTRALPGAWKQSGAGCRNNSG 360

QY 361 FPSNPKFWLRVSEPVYIAVLQSRHLHAADWAGARALVGDSTSWSPASIPCKHYQAV 420
DB 361 FPSNPKFWLRVSEPVYIAVLQSRHLHAADWAGARALVGDSTSWSPASIPCKHYQAV 420

QY 421 GLHLWKVKRRVNLPRVLSMPVAGTACHAYDREHVLRCLESPGYIYLAVPSTFLKDPAGE 480
DB 421 GLHLWKVKRRVNLPRVLSMPVAGTACHAYDREHVLRCLESPGYIYLAVPSTFLKDPAGE 480

QY 481 FLLRVFTSGRVSLSAIRAVAKNTTP 505
DB 481 FLLRVFTSGRVSLSAIRAVAKNTTP 505

RESULT 6

US-09-422-869-10
; Sequence 10, Application US/09422869
; Patent No. 6235481
; GENERAL INFORMATION:
; APPLICANT: POLONSKY, KENNETH S.
; APPLICANT: HORIKAWA, YUKIO
; APPLICANT: ODA, NAOHISA
; APPLICANT: COX, NANCY J.
; APPLICANT: SREENAN, SEAMUS
; APPLICANT: ZHOU, YUN-PING
; APPLICANT: OTANI, KENICHI
; APPLICANT: HANIS, CRAIG L.
; APPLICANT: BELL, GRAEME I.
; TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
; FILE REFERENCE: ARCD:307
; CURRENT APPLICATION NUMBER: US/09/422,869
; CURRENT FILING DATE: 1999-10-21
; EARLIER APPLICATION NUMBER: 60/134,175
; EARLIER FILING DATE: 1999-05-13
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 444
; TYPE: PRT
; ORGANISM: Human
US-09-422-869-10

Query Match 64.0%; Score 2336; DB 4; Length 444;
Best Local Similarity 99.8%; Pred. No. 2.3e-226;
Matches 425; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRAGGATPARELFDAAFPAAADSSLCFCDLSTPLAQFREDITWRRPOEICATPRLFPDDP 60
DB 1 MRAGGATPARELFDAAFPAAADSSLCFCDLSTPLAQFREDITWRRPOEICATPRLFPDDP 60

QY 61 REGOVKGLGDCWFLCACAALQKSRHLQDVIPPGQPSWADOEYRGSTCRINQFGRW 120
DB 61 REGOVKGLGDCWFLCACAALQKSRHLQDVIPPGQPSWADOEYRGSTCRINQFGRW 120

QY 121 EVTTDDRLPCLAGRLCFRCQREDVFWLPLEKYYAKVHGSYEHLMWAGOVADALVDLTGG 180
DB 121 EVTTDDRLPCLAGRLCFRCQREDVFWLPLEKYYAKVHGSYEHLMWAGOVADALVDLTGG 180

QY 181 LAERNLKGAVSGGQODRPGWEHRTCRQLHLKQDCLISCCVLSPRAGARELGEHAF 240
DB 181 LAERNLKGAVSGGQODRPGWEHRTCRQLHLKQDCLISCCVLSPRAGARELGEHAF 240

QY 241 IVSDRLQOGAGOCILLRIONPWRCWQGLWREGGEGSWQDAVAASELLSLOQEGE 300
DB 241 IVSDRLQOGAGOCILLRIONPWRCWQGLWREGGEGSWQDAVAASELLSLOQEGE 300

QY 301 FWEEEFRLREFDELTVGYPVTEAGHLQSLYTERLLCHTRALPGAWKQSGAGCRNNSG 360
DB 301 FWEEEFRLREFDELTVGYPVTEAGHLQSLYTERLLCHTRALPGAWKQSGAGCRNNSG 360

QY 361 FPSNPKFWLRVSEPVYIAVLQSRHLHAADWAGARALVGDSTSWSPASIPCKHYQAV 420
DB 361 FPSNPKFWLRVSEPVYIAVLQSRHLHAADWAGARALVGDSTSWSPASIPCKHYQAV 420

QY 421 GLHLWK 426
DB 421 GLHLWK 426

RESULT 7

US-09-422-869-12
; Sequence 12, Application US/09422869
; Patent No. 6235481
; GENERAL INFORMATION:
; APPLICANT: POLONSKY, KENNETH S.
; APPLICANT: HORIKAWA, YUKIO
; APPLICANT: ODA, NAOHISA
; APPLICANT: COX, NANCY J.
; APPLICANT: SREENAN, SEAMUS
; APPLICANT: ZHOU, YUN-PING
; APPLICANT: OTANI, KENICHI
; APPLICANT: HANIS, CRAIG L.
; APPLICANT: BELL, GRAEME I.
; TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
; FILE REFERENCE: ARCD:307
; CURRENT APPLICATION NUMBER: US/09/422,869
; CURRENT FILING DATE: 1999-10-21
; EARLIER APPLICATION NUMBER: 60/134,175
; EARLIER FILING DATE: 1999-05-13
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 274
; TYPE: PRT
; ORGANISM: Human
US-09-422-869-12

Query Match 23.5%; Score 858.5; DB 4; Length 274;
Best Local Similarity 81.7%; Pred. No. 4.6e-78;
Matches 165; Conservative 1; Mismatches 13; Indels 23; Gaps 3;

QY 1 MRAGGATPARELFDAAFPAAADSSLCFCDLSTPLAQFREDITWRRPOEICATPRLFPDDP 60
DB 1 MRAGGATPARELFDAAFPAAADSSLCFCDLSTPLAQFREDITWRRPOEICATPRLFPDDP 60

QY 61 RESOVKGLIGDCWFLCACAALOKSRHLLDVIIPGQPSWADQETRGSTCRIMQGRNV 120
 DB 61 RESOVKGLIGDCWFLCACAALOKSRHLLDVIIPGQPSWADQETRGSTCRIMQGRNV 120
 QY 121 EYTTDRLECLAGRCFSCQREDDVFWPLLEKVKYAKVGSYEHLMAGVADALVDLTG 180
 DB 121 EYTTDRLECLAGRCFSCQREDDVFWPLLEKVKYAKVGSYEHLMAGVADALVDLTG 180
 QY 181 LAERNMLKGVAGSGGQDDRRGR 202
 DB 168 -----GCPGGRDRRGR 179

RESULT 8

US-08-835-099A-1
 ; Sequence 1, Application US/08835099A
 ; Patent No. 5874277
 ; GENERAL INFORMATION:
 ; APPLICANT: SHINTANI, Yasushi
 ; APPLICANT: NISHI, Kazuo
 ; APPLICANT: KAWAMOTO, Tomohiro
 ; TITLE OF INVENTION: NOVEL PROTEINS, THEIR PRODUCTION
 ; NUMBER OF SEQUENCES: 18
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP
 ; STREET: 130 Water Street
 ; CITY: Boston
 ; STATE: MA
 ; COUNTRY: USA
 ; ZIP: 02109
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FASTSEQ for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/835, 099A
 ; FILING DATE: 04-APR-1997
 ; CLASSIFICATION: 514
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 083649/1996
 ; FILING DATE: 05-APR-1996
 ; APPLICATION NUMBER: 97105508.2
 ; FILING DATE: 03-APR-1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Resnick, David S
 ; REGISTRATION NUMBER: 34,235
 ; REFERENCE/DOCKET NUMBER: 47342
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 617-523-3400
 ; TELEFAX: 617-523-6440
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 703 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-835-099A-1

Query Match 17.4%; Score 635; DB 2; Length 703;
 Best local Similarity 32.3%; Pred. No. 6, 5e-55;
 Matches 164; Conservative 85; Mismatches 179; Indels 80; Gaps 17;

QY 13 LFRDAFPADSSL-FCDSLSTPLAQFREDITWRROEICATPRLFPDDPREGVKGLG 71
 DB 45 LFRDPPAPCPASALGYKDLG-PCSPOTQGIIMKPTLCSPOFVIGAGARTDLCQGLG 103
 QY 72 DCNFIACALQKSRHLLDVIIPGQPSWADQETRGSTCRIMQGRNV 131

DB 104 DCMILAATASLTINLELLIYRVPRD--FOENTAGITHFQWQYGEVVEVYIDRLPTK 161
 QY 132 AGRCLFSCQREDDVFWPLLEKVKYAKVGSYEHLMAGVADALVDLTGLAERNMLKGA 191
 DB 162 NGQLFLHSEQGENEFMSALLEKAYAKINGCYEALAGSTVGEFEDFTGISFYLK--- 218
 QY 192 GSGGQDDRRGRHRTCCQLHLKQC---LISCCV-LSPRAGA-----RELGEFHATV 242
 DB 219 -----KPNALYQIIRKAL-----CAGSLGCSIDVSAAEATISOKLYSHASY 266
 QY 243 SDRLEIQAGCCILLIRIOWPMGRQWGLMREGEGMSOVDAVAASELLSOLEGFEW 302
 DB 267 TGVEEVNQ-GHEPLILRNPMGVEBMSGAMSDAPFNHIDPRKKEELKVEDGEFW 325
 QY 303 VEEEFLEPDELIVGYVTENAGHLQSLYTERLLCHTRALPGAWYKGSAGCRRNSGFP 362
 DB 326 MSLDFVROFRLERLCNLSPLSLSEEVHKNMVLFN---GHWTRGSTAGCCQN---YP 378
 QY 363 ----SNPKFRLVSEPS-----YIYAVLQSRILHADMAGRALYDSTHS 406
 DB 379 ATYTNPOFKIRLDEVDDEESIGECCTVLLGIMQNR-----RMKRRIGQGLMS 430
 QY 407 WSPASIPGKHVQAVGLHMKYER-----RVNLP--VLSMPVAGTACHAYDREYHLRC 459
 DB 431 -----IGYAVYQPKLESHTDHLGRDFLAYQPSARTSTYVNLKEVSGRA 477
 QY 460 ELSPGYIYAVSTLKAAPGELLRVPS 487
 DB 478 RLPPGYIYVSPTEPEPGEFLRVPS 505

RESULT 9

US-09-157-349-1
 ; Sequence 1, Application US/09157349
 ; Patent No. 6068990
 ; GENERAL INFORMATION:
 ; APPLICANT: SHINTANI, Yasushi
 ; APPLICANT: NISHI, Kazuo
 ; APPLICANT: KAWAMOTO, Tomohiro
 ; TITLE OF INVENTION: NOVEL PROTEINS, THEIR PRODUCTION
 ; NUMBER OF SEQUENCES: 18
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP
 ; STREET: 130 Water Street
 ; CITY: Boston
 ; STATE: MA
 ; COUNTRY: USA
 ; ZIP: 02109
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FASTSEQ for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/157, 349
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/835, 099
 ; FILING DATE:
 ; APPLICATION NUMBER: 97105508.2
 ; FILING DATE: 03-APR-1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Resnick, David S
 ; REGISTRATION NUMBER: 34,235
 ; REFERENCE/DOCKET NUMBER: 47342
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 617-523-3400
 ; TELEFAX: 617-523-6440
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 703 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-157-349-1

Query Match 17.4%; Score 635; DB 3; Length 703;
Best Local Similarity 32.3%; Pred. No. 6.5e-55;
Matches 164; Conservative 85; Mismatches 179; Indels 80; Gaps 17;

QY 13 LFRDAAPPAADSSL-FCDLSTPLAOFREDITWRRPOEICATPRLFPDDPREQVKGGLG 71
DB 45 LFKDPEFPACPSALGYKDLG-PCSPOTOGIWKRTCLPSPQFIVGGATRDICQGG 103
QY 72 DCWFLCACAAOKSHLLDQVIPPQPSWADQYEGSFTCRWQGRWVETDRLPCL 131
DB 104 DCWLLAAIASLNEELLYRVVPRDQ--FOENYAGIFHFQWYGEVVEVVDRLPTK 161
QY 132 AGRLCFSQCRQEDVFWLPLEKVKYAKVHGSYEHLMWAGOVADALVDLTGGLAERNLKGVA 191
DB 162 NGOLLFLHSEQGNFWSALLEKAYAKLNGCYEALAGGSTVEGFEDFTGGISEFYDLK--- 218
QY 192 GSGGQDRPRWEHRTCRQLLHLKQDQ---LISCCV-LSPRAGA-----RELGEFHAFIV 242
DB 219 -----KPPANLYQIIRKAL-----CAGSLGCSIDVSSAAEAFAITSQKLVKSHAYSV 266
QY 243 SDRLEQOAGQCILLRIQNPWGRRCWQGLWREGGEGSQVDAVASSELLSQLEGFW 302
DB 267 TGVVEVNFQ-GHPEKLIIRLNPNWGEVSGAWSDDAPEWNNHIDPRRKEELDKKVEDGEFW 325
QY 303 VEEEFLEFREFDELTVGYPVTEAGHLSQLYTERLLCHTRALPGAWYKSGAGCRNNSGFP 362
DB 326 MSLSDVFRQFSRLICNLSPDLSSEEVHKNLVLFN-----GHWTRGSTAGCQCN---YP 378
QY 363 -----SNPKFWLRVSEPE-----VYIAVLRSLRHAADWAGARALVGDSTHS 406
DB 379 ATYTNPFQKIRLDEVEDQESIGEPCTVLLGLMKNR-----RWRKRIGQGMLS 430
QY 407 WSPASIPKHYQAVGLHLWKVEKR-----RVNLPK--VLSMPPVAGTACHAYDREVHLRC 459
DB 431 -----ICGAYVQPKLESHTDHGLGRDFFLAYQPSARTSTYVNLREVSGRA 477
QY 460 ELSPGYLVAVPSTFLKDPAGEFLLRVFS 487
DB 478 RLPPGEYLVPSTFEPPKDGECFLRVFS 505

RESULT 10
US-08-835-099A-2
Sequence 2, Application US/08835099A
Patent No. 5874277
GENERAL INFORMATION:
APPLICANT: SHINTANI, Yasushi
APPLICANT: NISHI, Kazuori
APPLICANT: KAWAMOTO, Tomohiro
TITLE OF INVENTION: NOVEL PROTEINS, THEIR PRODUCTION
TITLE OF INVENTION: AND USE
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/835.099A
FILING DATE: 04-APR-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 083649/1996
FILING DATE: 05-APR-1996
APPLICATION NUMBER: 97105508.2
FILING DATE: 03-APR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Resnick, David S
REGISTRATION NUMBER: 34,235
REFERENCE/DOCKET NUMBER: 47342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
TELEX:

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 712 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-835-099A-2

Query Match 17.4%; Score 635; DB 2; Length 712;
Best Local Similarity 32.3%; Pred. No. 6.6e-55;
Matches 164; Conservative 85; Mismatches 179; Indels 80; Gaps 17;

QY 13 LFRDAAPPAADSSL-FCDLSTPLAOFREDITWRRPOEICATPRLFPDDPREQVKGGLG 71
DB 54 LFKDPEFPACPSALGYKDLG-PCSPOTOGIWKRTCLPSPQFIVGGATRDICQGG 112
QY 72 DCWFLCACAAOKSHLLDQVIPPQPSWADQYEGSFTCRWQGRWVETDRLPCL 131
DB 113 DCWLLAAIASLNEELLYRVVPRDQ--FOENYAGIFHFQWYGEVVEVVDRLPTK 170
QY 132 AGRLCFSQCRQEDVFWLPLEKVKYAKVHGSYEHLMWAGOVADALVDLTGGLAERNLKGVA 191
DB 171 NGOLLFLHSEQGNFWSALLEKAYAKLNGCYEALAGGSTVEGFEDFTGGISEFYDLK--- 227
QY 192 GSGGQDRPRWEHRTCRQLLHLKQDQ---LISCCV-LSPRAGA-----RELGEFHAFIV 242
DB 228 -----KPPANLYQIIRKAL-----CAGSLGCSIDVSSAAEAFAITSQKLVKSHAYSV 275
QY 243 SDRLEQOAGQCILLRIQNPWGRRCWQGLWREGGEGSQVDAVASSELLSQLEGFW 302
DB 276 TGVVEVNFQ-GHPEKLIIRLNPNWGEVSGAWSDDAPEWNNHIDPRRKEELDKKVEDGEFW 334
QY 303 VEEEFLEFREFDELTVGYPVTEAGHLSQLYTERLLCHTRALPGAWYKSGAGCRNNSGFP 362
DB 335 MSLSDVFRQFSRLICNLSPDLSSEEVHKNLVLFN-----GHWTRGSTAGCQCN---YP 387
QY 363 -----SNPKFWLRVSEPE-----VYIAVLRSLRHAADWAGARALVGDSTHS 406
DB 388 ATYTNPFQKIRLDEVEDQESIGEPCTVLLGLMKNR-----RWRKRIGQGMLS 439
QY 407 WSPASIPKHYQAVGLHLWKVEKR-----RVNLPK--VLSMPPVAGTACHAYDREVHLRC 459
DB 440 -----ICGAYVQPKLESHTDHGLGRDFFLAYQPSARTSTYVNLREVSGRA 486
QY 460 ELSPGYLVAVPSTFLKDPAGEFLLRVFS 487
DB 487 RLPPGEYLVPSTFEPPKDGECFLRVFS 514

RESULT 11
US-09-157-349-2
Sequence 2, Application us/09157349
Patent No. 6068990
GENERAL INFORMATION:

Fri Sep 20 10:47:44 2002

us-09-768-877-2.ra1

Page 7

APPLICANT: SHINTANI, Yasushi
APPLICANT: NISHI, Kazuori
APPLICANT: KAWAMOTO, Tomohiro
TITLE OF INVENTION: NOVEL PROTEINS, THEIR PRODUCTION
TITLE OF INVENTION: AND USE
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESS: DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/157,349
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/835,099
FILING DATE:
APPLICATION NUMBER: 97105508.2
FILING DATE: 03-APR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Resnick, David S
REGISTRATION NUMBER: 34,235
REFERENCE/DOCKET NUMBER: 47342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 712 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Protein
US-09-157-349-2

Query Match 17.4% Score 635; DB 3; Length 712;
Best Local Similarity 32.3%; Pred. No. 6,6e-55;
Matches 164; Conservative 85; Mismatches 179; Indels 80; Gaps 17;
QY 13 LFRDAFPAADSSL-FCDSLTPLAQFREDITWRPQICATPRLFPDDPREGQKGLG 71
DB 54 LKDEFPAPCALGKDLG-PGSPOTGCIIMKRPTELCPFOIYVGATRDICGGG 112
QY 72 DCFWLCACAAALOKSRHLDOVIPPGQPSWADDEYRGSFTCRIMQFGRWVEYTTDRPLCL 131
DB 113 DCMWLAIASTLNLDELILRYVPRDOD--FOENYAGIFHFQFQMGVEWYVYDRLPLK 170
QY 132 AGRLECSRCORREDFVPLPLEKVAKVHGSYELHNAQYADLVLTGLAERMLKGYA 191
DB 171 NCQLFLHSEQNEFWSALEKAVAKLNGCYALAGSGTVECFEFTGGSISEFYDLK 227
QY 192 GSGGQODRPGRWEHRTCRQLHLKDKC--LISCCV-LSPRAGA-----RELGEFHAFTV 242
DB 228 -----KPRANLYQIIRKAL-----CAGSLGCSIDVSSAAEAETISQKLKSHAVSV 275
QY 243 SDRELQOAGCCILLRIQNPWGRRCMOGLMREGGEGNSQYDAVAASSELLSQLOEGEFWE 302
DB 276 TGVEEVNPO-GHPEKILIRLNPWGEVEMSGAAMDAPENHIDPRKRELDKKVDGEFWM 334
QY 303 VEEEFLEFDELIVGYPTENGHLSLYTERLCHTRALPGAWYKGSAGCCRNNSGFP 362
DB 335 MSLDFVHQSRLTICNTLSPELSSEEVHKNMLVLEF--GHWTGSGTAGCCCN---YF 387

QY 363 -----SNPKFLVSESE-----VYIAVLQSRSLHAADWAGARALVDSHTS 406
DB 388 ATWTNPQFKIRLDEVEDDEESIGEPCCVTLGLMOKNR-----RRKRIGQCMLS 439
QY 407 MSPASIPGKHVQAVGLHMKVER-----RVNLP--VLSMPYAGTACIAYREYHLC 459
DB 440 -----IGVAVVQPKELSHYDAHLGRDFLAVOPSARTSYVNLREVSRA 486
QY 460 ELSPGYLAVPSTFLKDAPEFLRVPS 487
DB 487 RLPPGEYLVVSTFEPPKDEFCFLRVPS 514
RESULT 12
US-09-422-869-22
Sequence 22, Application US/09422869
Patent No. 6235481
GENERAL INFORMATION:
APPLICANT: POLONSKY, KENNETH S.
APPLICANT: HORIKAWA, YUKIO
APPLICANT: ODA, NAOHISA
APPLICANT: COX, NANCY J.
APPLICANT: SEENAN, SEAMUS
APPLICANT: ZHOU, YUN-PING
APPLICANT: OTANI, KENICHI
APPLICANT: HANIS, CRAIG L.
APPLICANT: BELL, GRAEME I.
TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
FILE REFERENCE: ARCD:307
CURRENT APPLICATION NUMBER: US/09/422,869
CURRENT FILING DATE: 1999-10-21
EARLIER APPLICATION NUMBER: 60/134,175
EARLIER FILING DATE: 1999-05-13
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 22
LENGTH: 714
TYPE: PPT
ORGANISM: Human
US-09-422-869-22

Query Match 16.9% Score 618.5; DB 4; Length 714;
Best Local Similarity 32.3%; Pred. No. 3e-53;
Matches 166; Conservative 73; Mismatches 198; Indels 77; Gaps 17;

QY 13 LFRDAFPAADSSL-FCDSLTPLAQFREDITWRPQICATPRLFPDDPREGQKGLG 71
DB 55 LFRDEAFPPVPSISLGKDLG-FNSKTYGIMKRPTELSNQFIYDGAITRDICGALG 113
QY 72 DCFWLCACAAALOKSRHLDOVIPPGQPSWADDEYRGSFTCRIMQFGRWVEYTTDRPLCL 131
DB 114 DCMWLAIASTLNLDELILRYVPRDOD--FONGYAGIFHFQFQMGVEWYVYDRLPLK 171
QY 132 AGRLECSRCORREDFVPLPLEKVAKVHGSYELHNAQYADLVLTGLAERMLKGYA 191
DB 172 DKLIVFVHSAEGNEFWSALEKAVAKLNGCYALAGSGTSECFEFTGCVTEWEYELKAP 231
QY 192 GSGGQODRPGRWEHRTCRQLHLKDKC--LISCCV-LSPRAGA-----RELGEFHAFTV 242
DB 232 SDLYO-----LTKALERSLGCSDISSVDMETLFFKLVKHAHVSYGA 279
QY 246 RELQOAGCCILLRIQNPWGRRCMOGLMREGGEGNSQYDAVAASSELLSQLOEGEFWE 305
DB 280 KQV-NYRGQVSLIRBNPWGEVEMSGAAMDAPENHIDPRKRELDKKVDGEFWM 338
QY 306 EEFLEFDELIVGYPTENGHLSLYTERLCHTRALPGAWYKGSAGCCRNNSGFP 363
DB 339 RQFMEFTELICNTLTPDA--LKSFTIR--NNTLYEGTWRGSGTAGCCRN---YPAF 391
QY 364 --NPKFVLVSESE-----VYIAVLQSRSLHAADWAGARALVDSHTS 406
DB 392 WYNPQFKIRLDEVEDDEESIGEPCCVTLGLMOKNR-----RRKRIGQCMLS 439

us-09-768-877-2.ra

Page 9

RESULT 15
US-08-726-525-7
; Sequence 7, Application US/08726525

RESULT 15
 US-08-726-525-7
 : Sequence 7, Application US/08726525
 : Patent No. 5789181
 : GENERAL INFORMATION:
 : APPLICANT: Lih, Lih-Liang
 : APPLICANT: Graham, James
 : TITLE OF INVENTION: NOVEL INTERLEUKIN-1 RECEPTOR
 : TITLE OF INVENTION: INTRACELLULAR LIGAND PROTEINS AND INHIBITORS OF LIGAND
 : TITLE OF INVENTION: BINDING
 : NUMBER OF SEQUENCES: 7
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: LEGAL AFFAIRS, GENETICS INSTITUTE, INC.
 : STREET: 87 Cambridgepark Drive
 : CITY: Cambridge
 : STATE: MA
 : COUNTRY: USA
 : ZIP: 02140
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.25
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/726,525
 : FILING DATE: 07-OCT-1996
 : CLASSIFICATION: 435
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: 08/487,942
 : FILING DATE: 07-JUN-1995
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Brown, Scott A.
 : REGISTRATION NUMBER: 32,724
 : REFERENCE/DOCKET NUMBER: G15258
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (617) 498-8224
 : TELEFAX: (617) 876-5851
 : INFORMATION FOR SEQ ID NO: 7:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 700 amino acids
 : TYPE: amino acid
 : STRANDEDNESS: single
 : TOPOLOGY: linear
 : MOLECULE TYPE: protein
 : HYPOTHEICAL: NO
 : US-08-726-525-7

Query Match 16.6% Score 604; DB 1: Length 700;
Best Local Similarity 30.4% Pred. NO. 8.4e-52;
Matches 157; Conservative 76; Mismatches 209; Indels 74; Gaps 15.

```

Db      219  -----KPPMLFKIIOKA-----LQSGSLGCSIDITSAAMSEALTPOKIVKGNHVS  266
QY      243  SDLRELQOAOQCILLRIQWPNGRCHQOGIAREGCGEWSOVDAAVASELLSQLOGEFW  3020
Db      267  TGAEEVESN-GLQKLIIRINPWEVEEMTGMNANCPSMWNTIDPEERERLRTREHDEFW  3252
QY      303  VEEEEFLREFDELTAIVGVYVTEAGHQSILYTERLLCHP-----RALPGAWKOSQSGCR  3366
Db      326  MSFDFELRHRYL-----EICNLTPDITLSDYTKKMKLLTKQGNRRRSTAGGCR  3757
QY      357  N-NSGFPSPNPKFWLTVSEPS-----VYIAVLQSRSLHAADWAGRAPALVGSHT  4050
Db      376  NYPTFMWNPQYLILKLEBEDDEDEBDSGCTFLVGLQOKNR-----RRQRKMGEDMHT  4288
QY      406  -SMSBPASIPGCHYAQVGLHLK-----VEKRYVNLPRVLSMPRYAGTACHAYDREVHLRCE  4600
Db      429  IGFGEIYEPBELSGQOTIHLSKPNFLYNNRAERSDFTINL-----REVILNRFK  4766
QY      461  LSPGYVLAVPSTFLKADAGEFLIAVSESTGRVLSAI  496
Db      477  LPPGEYILVSTPEPNQDGCILINVESEKADYQAV  512

```

Search completed: September 10, 2002, 10:40:13
Job time: 327 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 19, 2002, 14:49:21 ; Search time 34.08 Seconds

(without alignments)
3411.168 Million cell updates/sec

Title: US-09-768-877-2

Perfect score: 3649

Sequence: 1 MRAGRGAPARELFRDAFP.....SOEMLGQFLQEVSVNAVMT 672

Scoring table: BLOSUM62

Searched: 562222 segs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SPTREMBL_19:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.podent:*
12: sp.virus:*
13: sp.vertibrate:*
14: sp.unclassified:*
15: sp.virus:*
16: sp.bacteriophage:*
17: sp.archaeal:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3649	100.0	672	4	09HC96
2	3649	100.0	705	4	096J12
3	3649	100.0	720	4	096J12
4	3383.5	92.7	653	6	095L14
5	2969	81.2	666	11	099J13
6	2969	81.2	666	11	099J13
7	2958	81.1	666	11	099J13
8	2715.5	74.4	517	4	09HC94
9	2701.5	74.0	544	4	09HC94
10	2699	74.0	513	4	09HC93
11	2691	73.7	605	11	09WVFO
12	2336	64.0	444	4	09HC92
13	1091	29.9	264	11	09HC92
14	858.5	23.5	274	4	09HC91
15	657.5	18.0	720	11	09ER56
16	647	17.7	702	13	0918T0

17	644	17.6	715	13	042133	042133 gallus gall
18	639	17.5	705	13	09YRC1	09YRC1 coturnix co
19	631	17.3	650	6	09N0M7	09N0M7 sus scrofa
20	628.5	17.2	716	6	09N185	09N185 bos taurus
21	624	17.1	778	5	09N4B1	09N4B1 caenorhabdi
22	617.5	16.9	640	11	091YU0	091YU0 mus musculu
23	617	16.9	714	6	09G1G2	09G1G2 macaca fasc
24	615.5	16.9	640	11	008688	008688 mus musculu
25	614.5	16.8	703	11	091V43	091V43 mus musculu
26	613.5	16.8	815	6	09G1G7	09G1G7 macaca fasc
27	612	16.8	703	11	064698	064698 ratu
28	611.5	16.8	790	5	09N4B2	09N4B2 caenorhabdi
29	609.5	16.7	709	6	09XSJ3	09XSJ3 oryctolagus
30	609	16.7	737	11	070376	070376 ratu
31	608.5	16.7	757	11	070482	070482 ratu
32	608.5	16.7	785	11	09Q2F9	09Q2F9 ratu
33	607	16.6	700	6	09HBB1	09HBB1 homo sapien
34	607	16.6	700	6	09GLG1	09GLG1 macaca fasc
35	606.5	16.6	709	11	008702	008702 ratu
36	604.5	16.6	704	13	09T8G2	09T8G2 brachydanio
37	597	16.4	822	6	09TTH9	09TTH9 bos taurus
38	594.5	16.3	634	4	000263	000263 mus musculu
39	594.5	16.3	709	11	088977	088977 mus musculu
40	593.5	16.3	821	6	046596	046596 sus scrofa
41	592	16.2	502	11	09ER55	09ER55 mus musculu
42	589.5	16.2	709	6	09XSJ1	09XSJ1 bos taurus
43	589	16.1	822	6	09TTH8	09TTH8 ovis aries
44	585	16.0	565	5	09V8U6	09V8U6 drosophila
45	583.5	16.0	709	6	09XSJ2	09XSJ2 sus scrofa

ALIGNMENTS

RESULT 1
ID 09HC96 PRELIMINARY; PRT; 672 AA.
AC 09HC96;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE TYPE A CALPAIN-LIKE PROTEASE (CALPAIN 10).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed-11017071;
RA Horikawa Y., Oda N., Cox N.J., Li X., Orho-Melander M., Hara M.,
RA Horioka Y., Lindner T.H., Mashima H., Schwarz P., del Bosque Plata L.,
RA Horikawa Y., Oda Y., Yoshituchi I., Collila S., Polonsky K.S., Wei S.,
RA Concannon P., Iwasaki N., Schulze J., Baier L.J., Bogardus C.,
RA Groop L., Boerwinkle E., Hants C.L., Bell G.I.,
RA "Genetic variation in the gene encoding calpain-10 is associated with
RT type 2 diabetes mellitus."
RL Nat. Genet. 26:163-175(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX PubMed-11017071;
RA Strassberg R.;
RC TISSUE-ENDOMETRIAL ADENOCARCINOMA;
RA Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL; AF089088; AAG17966.1; -;
DR EMBL; BC004260; AAH04260.1; -;
DR HSSP; Q07009; IDFO;
DR MEROPS; C02.018; -;
DR InterPro: IPR001300; Peptidase_C2.
DR InterPro: IPR001659; Thiolprol_act_site.
DR Pfam: PF01067; Calpain_III; 2.
DR Pfam: PF00648; Peptidase_C2; 1.
DR PRINTS; PR00704; CALPAIN;
DR SMART; SM00230; Cyspc; 1.
DR PROSITE; PS00139; THIOL_PROTEASE_CYS; UNKNOWN_1.

KW Protease.
SQ SEQUENCE 672 AA; 74938 MW; 74A48D879E997EE1 CRC64;

Query Match 100.0%; Score 3649; DB 4; Length 672;
Best Local Similarity 100.0%; Pred. No. 8e-298;
Matches 672; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRAGRGATPARELFRDAAFPAADSSFLCDLSTPLAQFREDITWRPQIEICATPRFLFDDP 60
Db 1 MRAGRGATPARELFRDAAFPAADSSFLCDLSTPLAQFREDITWRPQIEICATPRFLFDDP 60

QY 61 REGOVKGLGDCWFLCACAAALOKSRHLLDQVIPPQOPSWADQYRGSTFCRIWQGRWV 120
Db 61 REGOVKGLGDCWFLCACAAALOKSRHLLDQVIPPQOPSWADQYRGSTFCRIWQGRWV 120

QY 121 EVTTDDRLPCLAGRLCFRCQREDVFWLPLEKVIYAKVHGSYEHLMWAGOVADALVDLTGG 180
Db 121 EVTTDDRLPCLAGRLCFRCQREDVFWLPLEKVIYAKVHGSYEHLMWAGOVADALVDLTGG 180

QY 181 LAERNWLVAGSGGQODRPGWEHRTCRQLLHLKDQCLISCCVLSPPRAGARELGEHAF 240
Db 181 LAERNWLVAGSGGQODRPGWEHRTCRQLLHLKDQCLISCCVLSPPRAGARELGEHAF 240

QY 241 IVSDRLRLOQAGQCILLRIQNPWGRRCWQGLWREGGEGWSQVDAVAASELLSQLOE 300
Db 241 IVSDRLRLOQAGQCILLRIQNPWGRRCWQGLWREGGEGWSQVDAVAASELLSQLOE 300

QY 301 FWVEEEFLREFDELTVGYPTVTEAGHLQSLYTERLLCHTRALPGAWYKVGOSAGGCRNNSG 360
Db 301 FWVEEEFLREFDELTVGYPTVTEAGHLQSLYTERLLCHTRALPGAWYKVGOSAGGCRNNSG 360

QY 361 FPSNPFWLRVSEPVSEVYIAVLQSRHLHAADWAGRALYGDSTSWSPASIPGKHQYQAV 420
Db 361 FPSNPFWLRVSEPVSEVYIAVLQSRHLHAADWAGRALYGDSTSWSPASIPGKHQYQAV 420

QY 421 GLHLWKEVRKRVNLPVLSMPVAGTACHAYDREVLHRCLESPGYILAVPSTFLKDAPGE 480
Db 421 GLHLWKEVRKRVNLPVLSMPVAGTACHAYDREVLHRCLESPGYILAVPSTFLKDAPGE 480

QY 481 FLRLVSTGRVLSAIAVAKNTTPGAALPAGEWTVQLRGSWRVGTAGGSRNFSYPT 540
Db 481 FLRLVSTGRVLSAIAVAKNTTPGAALPAGEWTVQLRGSWRVGTAGGSRNFSYPT 540

QY 541 NPCFPSPVPEGPGRVITLHQHCRPSDTEFHPHIGFHFQVPEGGRSQDAPPLLLQLEPL 600
Db 541 NPCFPSPVPEGPGRVITLHQHCRPSDTEFHPHIGFHFQVPEGGRSQDAPPLLLQLEPL 600

QY 601 LSCVPHRYAQEVSRLCLLPAGTYKVVPSTYLPDTEGAFVTIATRIDRPSIHSEMLGQF 660
Db 601 LSCVPHRYAQEVSRLCLLPAGTYKVVPSTYLPDTEGAFVTIATRIDRPSIHSEMLGQF 660

QY 661 LQEVSYMAVYKMT 672
Db 661 LQEVSYMAVYKMT 672

RESULT 2
Q96J12 PRELIMINARY; PRT; 705 AA.
ID Q96J12
AC Q96J12
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DE KIAA1845 PROTEIN (FRAGMENT).
GN KIAA1845
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN NCBI_TaxID=9606;
RP [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;

MDLINE=21245130; PubMed=11347906;
RA Nagase T., Nakayama M., Nakajima D., Kikuno R., Ohara O.;
RT *Prediction of the coding sequences of unidentified human genes. XX.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.*;
RL DNA Res. 8:95-95(2001).
DR ENBL; AB058748; BAB47474.1; -.
FT NON_TER 1
SQ SEQUENCE 705 AA; 78287 MW; 3B3BA9824FDF0F53F CRC64;

Query Match 100.0%; Score 3649; DB 4; Length 705;
Best Local Similarity 100.0%; Pred. No. 8.6e-298;
Matches 672; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRAGRGATPARELFRDAAFPAADSSFLCDLSTPLAQFREDITWRPQIEICATPRFLFDDP 60
Db 34 MRAGRGATPARELFRDAAFPAADSSFLCDLSTPLAQFREDITWRPQIEICATPRFLFDDP 93

QY 61 REGOVKGLGDCWFLCACAAALOKSRHLLDQVIPPQOPSWADQYRGSTFCRIWQGRWV 120
Db 94 REGOVKGLGDCWFLCACAAALOKSRHLLDQVIPPQOPSWADQYRGSTFCRIWQGRWV 153

QY 121 EVTTDDRLPCLAGRLCFRCQREDVFWLPLEKVIYAKVHGSYEHLMWAGOVADALVDLTGG 180
Db 154 EVTTDDRLPCLAGRLCFRCQREDVFWLPLEKVIYAKVHGSYEHLMWAGOVADALVDLTGG 213

QY 181 LAERNWLVAGSGGQODRPGWEHRTCRQLLHLKDQCLISCCVLSPPRAGARELGEHAF 240
Db 214 LAERNWLVAGSGGQODRPGWEHRTCRQLLHLKDQCLISCCVLSPPRAGARELGEHAF 273

QY 241 IVSDRLRLOQAGQCILLRIQNPWGRRCWQGLWREGGEGWSQVDAVAASELLSQLOE 300
Db 274 IVSDRLRLOQAGQCILLRIQNPWGRRCWQGLWREGGEGWSQVDAVAASELLSQLOE 333

QY 301 FWVEEEFLREFDELTVGYPTVTEAGHLQSLYTERLLCHTRALPGAWYKVGOSAGGCRNNSG 360
Db 334 FWVEEEFLREFDELTVGYPTVTEAGHLQSLYTERLLCHTRALPGAWYKVGOSAGGCRNNSG 393

QY 361 FPSNPFWLRVSEPVSEVYIAVLQSRHLHAADWAGRALYGDSTSWSPASIPGKHQYQAV 420
Db 394 FPSNPFWLRVSEPVSEVYIAVLQSRHLHAADWAGRALYGDSTSWSPASIPGKHQYQAV 453

QY 421 GLHLWKEVRKRVNLPVLSMPVAGTACHAYDREVLHRCLESPGYILAVPSTFLKDAPGE 480
Db 454 GLHLWKEVRKRVNLPVLSMPVAGTACHAYDREVLHRCLESPGYILAVPSTFLKDAPGE 513

QY 481 FLRLVSTGRVLSAIAVAKNTTPGAALPAGEWTVQLRGSWRVGTAGGSRNFSYPT 540
Db 514 FLRLVSTGRVLSAIAVAKNTTPGAALPAGEWTVQLRGSWRVGTAGGSRNFSYPT 573

QY 541 NPCFPSPVPEGPGRVITLHQHCRPSDTEFHPHIGFHFQVPEGGRSQDAPPLLLQLEPL 600
Db 574 NPCFPSPVPEGPGRVITLHQHCRPSDTEFHPHIGFHFQVPEGGRSQDAPPLLLQLEPL 633

QY 601 LSCVPHRYAQEVSRLCLLPAGTYKVVPSTYLPDTEGAFVTIATRIDRPSIHSEMLGQF 660
Db 634 LSCVPHRYAQEVSRLCLLPAGTYKVVPSTYLPDTEGAFVTIATRIDRPSIHSEMLGQF 693

QY 661 LQEVSYMAVYKMT 672
Db 694 LQEVSYMAVYKMT 705

RESULT 3
Q96IG4 PRELIMINARY; PRT; 720 AA.
ID Q96IG4
AC Q96IG4;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE UNKNOWN (PROTEIN FOR IMAGE:2967328) (FRAGMENT).
OS Homo sapiens (Human).

OC Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC TISSUE-MELANOMA;
 RC Strausberg R.;
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC007553; AA07553.1; -
 FT NON-TER
 SQ SEQUENCE 720 AA; 79586 MW; 25CD98031078EFD6 CRC64;

Query Match 100.0%; Score 3649; DB 4; Length 720;
 Best Local Similarity 100.0%; Pred. No. 8.8e-298;
 Matches 672; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRAGGATPARELPDAAFPADSSLFCDLSTPLAQFREDITWRPPOETCATPRLFPDP 60
 DB 49 MRAGGATPARELPDAAFPADSSLFCDLSTPLAQFREDITWRPPOETCATPRLFPDP 108
 QY 61 REGGVKGLDGCWFLCACAALQKSRHLLDQVTPPGQSMADQERGSFTCRIMPGHW 120
 DB 109 REGGVKGLDGCWFLCACAALQKSRHLLDQVTPPGQSMADQERGSFTCRIMPGHW 168
 QY 121 EVTTDDRLPLCLAGRLCFRCQREDFVPLLEKYAKYHGSYEHLMAGQVADALVDLTGG 180
 DB 169 EVTTDDRLPLCLAGRLCFRCQREDFVPLLEKYAKYHGSYEHLMAGQVADALVDLTGG 228
 QY 181 LAERNMLKGVAGSGGQDPRGMEHRTCRQLHLKDCCLISCVLSPRAGARELGEFNAF 240
 DB 229 LAERNMLKGVAGSGGQDPRGMEHRTCRQLHLKDCCLISCVLSPRAGARELGEFNAF 288
 QY 241 IVSDLRELQAGOCILLRIQNPWGRRCWGLMEGEGMSQVDAVAASELLSOLQGE 300
 DB 289 IVSDLRELQAGOCILLRIQNPWGRRCWGLMEGEGMSQVDAVAASELLSOLQGE 348
 QY 301 FWVEEELREFDELTVGYPTTEAGHLSLYTERILCHTRALPGAMVWGQAGGCRNNSG 360
 DB 349 FWVEEELREFDELTVGYPTTEAGHLSLYTERILCHTRALPGAMVWGQAGGCRNNSG 408
 QY 361 FFSNPKFELRVSESEYIYIAVQSRRLAADMAGARALVDSHTSMPASIPGKHQAV 420
 DB 409 FFSNPKFELRVSESEYIYIAVQSRRLAADMAGARALVDSHTSMPASIPGKHQAV 468
 QY 421 GLHLMKVEKRRVNLPRVLSMPVAGTACHADREHNLCELSPGYLLAVPSTFLKDAGE 480
 DB 469 GLHLMKVEKRRVNLPRVLSMPVAGTACHADREHNLCELSPGYLLAVPSTFLKDAGE 528
 QY 481 FLRVFSTGRVSLSAITAVAKNTTPGAALPAGENGTVOLRGSWVGOTAGSRRNFASPT 540
 DB 529 FLRVFSTGRVSLSAITAVAKNTTPGAALPAGENGTVOLRGSWVGOTAGSRRNFASPT 588
 QY 541 NPCPFVSVPBGPRCVRITLHQCRPSDVEFHPIGFHIFQVPGSGSODAPRLLOEPL 600
 DB 589 NPCPFVSVPBGPRCVRITLHQCRPSDVEFHPIGFHIFQVPGSGSODAPRLLOEPL 648
 QY 601 LSCVPHRYADEVSRILCLLPAGTYKVPSTYLPDEGAFVTIATRIDRPSIHSOEMLGOF 660
 DB 649 LSCVPHRYADEVSRILCLLPAGTYKVPSTYLPDEGAFVTIATRIDRPSIHSOEMLGOF 708
 QY 661 LOEYSVAWAKKT 672
 DB 709 LOEYSVAWAKKT 720

RESULT 4
 ID 095LP4 PRELIMINARY; PRT; 653 AA.
 AC 095LP4;
 DT 01-DEC-2001 (T-EMBLrel. 19, Created)
 DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)

DE HYPOTHETICAL 73.0 KDA PROTEIN.
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecoidea; Macaca.
 NCBI_TaxID=9541;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC TISSUE-TESTIS;
 RC Hashimoto K., Osada N., Kusuda J., Tanuma R., Hirai M.,
 RA Teruo K., Sugano S.;
 RT "Isolation of novel full-length cDNA clones from macaque testis cDNA
 RT libraries";
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB02744; BAB69713.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 653 AA; 72397 MW; 47812497BB315971 CRC64;

Query Match 92.7%; Score 3383.5; DB 6; Length 653;
 Best Local Similarity 95.5%; Pred. No. 1.5e-275;
 Matches 622; Conservative 12; Mismatches 16; Indels 1; Gaps 1;

QY 1 MRAGGATPARELPDAAFPADSSLFCDLSTPLAQFREDITWRPPOETCATPRLFPDP 60
 DB 1 MRAGGATPARELPDAAFPADSSLFCDLSTPLAQFREDITWRPPOETCATPRLFPDP 60
 QY 61 REGGVKGLDGCWFLCACAALQKSRHLLDQVTPPGQSMADQERGSFTCRIMPGHW 120
 DB 61 REGGVKGLDGCWFLCACAALQKSRHLLDQVTPPGQSMADQERGSFTCRIMPGHW 120
 QY 121 EVTTDDRLPLCLAGRLCFRCQREDFVPLLEKYAKYHGSYEHLMAGQVADALVDLTGG 180
 DB 121 EVTTDDRLPLCLAGRLCFRCQREDFVPLLEKYAKYHGSYEHLMAGQVADALVDLTGG 180
 QY 181 LAERNMLKGVAGSGGQDPRGMEHRTCRQLHLKDCCLISCVLSPRAGARELGEFNAF 240
 DB 181 LAERNMLKGVAGSGGQDPRGMEHRTCRQLHLKDCCLISCVLSPRAGARELGEFNAF 240
 QY 241 IVSDLRELQAGOCILLRIQNPWGRRCWGLMEGEGMSQVDAVAASELLSOLQGE 300
 DB 241 IVSDLRELQAGOCILLRIQNPWGRRCWGLMEGEGMSQVDAVAASELLSOLQGE 300
 QY 301 FWVEEELREFDELTVGYPTTEAGHLSLYTERILCHTRALPGAMVWGQAGGCRNNSG 360
 DB 301 FWVEEELREFDELTVGYPTTEAGHLSLYTERILCHTRALPGAMVWGQAGGCRNNSG 360
 QY 361 FFSNPKFELRVSESEYIYIAVQSRRLAADMAGARALVDSHTSMPASIPGKHQAV 420
 DB 361 FFSNPKFELRVSESEYIYIAVQSRRLAADMAGARALVDSHTSMPASIPGKHQAV 420
 QY 421 GLHLMKVEKRRVNLPRVLSMPVAGTACHADREHNLCELSPGYLLAVPSTFLKDAGE 480
 DB 421 GLHLMKVEKRRVNLPRVLSMPVAGTACHADREHNLCELSPGYLLAVPSTFLKDAGE 480
 QY 481 FLRVFSTGRVSLSAITAVAKNTTPGAALPAGENGTVOLRGSWVGOTAGSRRNFASPT 540
 DB 481 FLRVFSTGRVSLSAITAVAKNTTPGAALPAGENGTVOLRGSWVGOTAGSRRNFASPT 540
 QY 541 NPCPFVSVPBGPRCVRITLHQCRPSDVEFHPIGFHIFQVPGSGSODAPRLLOEPL 600
 DB 541 NPCPFVSVPBGPRCVRITLHQCRPSDVEFHPIGFHIFQVPGSGSODAPRLLOEPL 600
 QY 601 LSCVPHRYADEVSRILCLLPAGTYKVPSTYLPDEGAFVTIATRIDRPSIHSOEMLGOF 650
 DB 601 LSCVPHRYADEVSRILCLLPAGTYKVPSTYLPDEGAFVTIATRIDRPSIHSOEMLGOF 651

RESULT 5
 ID 099J13 PRELIMINARY; PRT; 666 AA.
 AC 099J13;
 DT 01-JUN-2001 (T-EMBLrel. 17, Created)

01-JUN-2001 (Tremblrel. 17, Last sequence update)
01-DEC-2001 (Tremblrel. 19, Last annotation update)
CALPAIN 10.
CAPN10.
Mus musculus (Mouse).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
Strausberg R.;
Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
[2]
SEQUENCE FROM N.A.
TISSUE=LIVER, NORMAL. 5 MONTH OLD MALE MOUSE.;
Strausberg R.;
Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
EMBL; BC005681; AAH05681.1; -;
EMBL; BC010969; AAH10969.1; -;
HSP; Q07009; IDP0.
MEROPS; C02.018; -;
MGD; MGI:1344392; Capn10.
InterPro; IPR001300; Peptidase_C2.
InterPro; IPR001169; Thiolprot_act_site.
Pfam; PF01067; Calpain_III; 2.
Pfam; PF00648; Peptidase_C2; 1.
PRINTS; PR00704; CALPAIN.
SMART; SM00230; Cyspc; 1.
PROSITE; PS00139; THIOI_PROTEASE_CYS; UNKNOWN_1.
SEQUENCE 666 AA; 74596 MW; AOB1B37E1D2371 CRC64;

Query Match 81.4%; Score 2969; DB 11; Length 666;
Best Local Similarity 81.2%; Pred. No. 9.8e-241;
Matches 545; Conservative 45; Mismatches 75; Indels 6; Gaps 2;

QY 1 MRAGGATPARELFDRAAFPAADSLFCDLSTPLAQFREDITWRPQICATPLPDDP 60
DB 1 MRVRAETPARELFDRAAFPAADSLFCDLSTPLAQFREDITWRPQICATPLPDDP 60
QY 61 REGOVKGLGDCWFLCACAALQKSHLLDQVIPPQSPWADQYRGSTCRINQFGRW 120
DB 61 WEGQVKGLGDCWFLCACAALQKSHLLDQVIPPQSPWADQYRGSTCRINQFGRW 120
QY 121 EVTTDDRLPCLAGLRCFSRCQREDVFWLPLEKVAHVHAGVADALVDTGG 180
DB 121 EVTTDDRLPCLAGLRCFSRCQREDVFWLPLEKVAHVHAGVADALVDTGG 180
QY 181 LAERNLKVAGSGGQDPRGWEHRTCRQLLHLKDCCLISCCVLSPRAGARELGEHAF 240
DB 181 LAERNLKVAGSGGQDPRGWEHRTCRQLLHLKDCCLISCCVLSPRAGARELGEHAF 240
QY 241 IVSDRLQAGQAGCIIILLRIQNPWRRRCWGLWREGGEGSWQYDAVAASSELLQAGE 300
DB 241 IVSDRLQAGQAGCIIILLRIQNPWRRRCWGLWREGGEGSWQYDAVAASSELLQAGE 300

DB 535 NPCLPFSYPEGAGPRYIRITLQOHCRLSDSQLHPGFHFVQVADGENQDACSLILQPL 594
QY 601 LSCVPHRYAQEVSRCLLIPAGTYKVVPSYLPDPTGATVTIATRIDRPSHSQEMLCQF 660
DB 595 LSCVPHRYAQEVSRCLLIPAGTYKVVPSYLPDPTGATVTIATRIDRPSHSQEMLCQF 654
QY 661 LQEVSMVAVMK 671
DB 655 LQEVSMVAVMK 665

RESULT 6
Q9ESK3
ID Q9ESK3 PRELIMINARY; PRT; 666 AA.
AC Q9ESK3
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE CALPAIN-LIKE PROTEASE.
GN CAPN10.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20472315; PubMed=11017071;
RA Horikawa Y., Oda N., Cox N.J., Li X., Orho-Melander M., Hara M.,
RA Hinokio Y., Lindner T.H., Mashima H., Schwarz P.E.H.,
RA del Bosque-Platawa L., Horikawa Y., Oda Y., Yoshiuchi I., Collila S.,
RA Polonsky K.S., Wei S., Concannon P., Iwasaki N., Schulze J.,
RA Baier L.J., Bogardus C., Groop L., Boerwinkle E., Hanis C.L.,
RA Bell G.I.;
RT *Genetic variation in the calpain 10 gene (CAPN10) is associated with
RT type 2 diabetes.*;
RL Nat. Genet. 26:163-175(2000).
DR EMBL; AF089089; AAG17967.1; -;
DR HSP; Q07009; IDP0.
DR MEROPS; C02.018; -;
DR MGD; MGI:1344392; Capn10.
DR InterPro; IPR001300; Peptidase_C2.
DR InterPro; IPR001169; Thiolprot_act_site.
DR Pfam; PF01067; Calpain_III; 2.
DR Pfam; PF00648; Peptidase_C2; 1.
DR PRINTS; PR00704; CALPAIN.
DR SMART; SM00230; Cyspc; 1.
DR PROSITE; PS00139; THIOI_PROTEASE_CYS; UNKNOWN_1.
KW Protease.
SQ SEQUENCE 666 AA; 74656 MW; BCC8D6778DA74671 CRC64;

Query Match 81.2%; Score 2963; DB 11; Length 666;
Best Local Similarity 81.1%; Pred. No. 3.1e-240;
Matches 544; Conservative 45; Mismatches 76; Indels 6; Gaps 2;

QY 1 MRAGGATPARELFDRAAFPAADSLFCDLSTPLAQFREDITWRPQICATPLPDDP 60
DB 1 MRVRAETPARELFDRAAFPAADSLFCDLSTPLAQFREDITWRPQICATPLPDDP 60
QY 61 REGOVKGLGDCWFLCACAALQKSHLLDQVIPPQSPWADQYRGSTCRINQFGRW 120
DB 61 WEGQVKGLGDCWFLCACAALQKSHLLDQVIPPQSPWADQYRGSTCRINQFGRW 120
QY 121 EVTTDDRLPCLAGLRCFSRCQREDVFWLPLEKVAHVHAGVADALVDTGG 180
DB 121 EVTTDDRLPCLAGLRCFSRCQREDVFWLPLEKVAHVHAGVADALVDTGG 180
QY 181 LAERNLKVAGSGGQDPRGWEHRTCRQLLHLKDCCLISCCVLSPRAGARELGEHAF 240
DB 181 LAERNLKVAGSGGQDPRGWEHRTCRQLLHLKDCCLISCCVLSPRAGARELGEHAF 240
QY 241 IVSDRLQAGQAGCIIILLRIQNPWRRRCWGLWREGGEGSWQYDAVAASSELLQAGE 300
DB 241 IVSDRLQAGQAGCIIILLRIQNPWRRRCWGLWREGGEGSWQYDAVAASSELLQAGE 300

```

Db 241 IISDLQELNSQOGILLRLRHNPWGRRCMOGLMRBGSGMNOVEPAKESSELLAQOEGE 300
Oy 301 FWESEEFRLREDELTVGYPVTEAGHLSLYTERLLCHTRALPGANVKQSGAGCCRNNSG 360
Db 301 FWESEEFRLREDELTVGYPVTEAGHLSLYTERLLCHTRALPGANVKQSGAGCCRNNSG 360
Oy 361 FPNKPFMLRLPESEVCAVYQRRRL---VGTALAG---ASPAVNLPGKYOAV 420
Db 361 FPNKPFMLRLPESEVCAVYQRRRL---VGTALAG---ASPAVNLPGKYOAV 414
Oy 421 GLHMKVEKRRVNLPRVLSMPVAGTACHAYDREYHLRCELSPGYLLAVSTFLKDAPGE 480
Db 421 GLHMKVEKRRVNLPRVLSMPVAGTACHAYDREYHLRCELSPGYLLAVSTFLKDAPGE 474
Oy 481 FLIRVSTGRVSLSAIRAVAKNTPGALPAEMGTVOGLRGSRVQGTAGSRNASTYPT 540
Db 475 FLIRVSTGRVSLSAIRAVAKNTPGALPAEMGTVOGLRGSRVQGTAGSRNASTYPT 534
Oy 541 NRCPEFVPEGPGRCVTRITLHQCPSDTEFFHPIGFHIFQYPEGGRSODAPILLQEBL 600
Db 535 NRCPEFVPEGPGRCVTRITLHQCPSDTEFFHPIGFHIFQYPEGGRSODAPILLQEBL 594
Oy 601 LSCVPHRYAEOEVSRLCLLPAGTYKVVPSTYLDTEGAFVTYATRIDRPSIHSOEMLAGOF 660
Db 595 LSCVPHRYAEOEVSRLCLLPAGTYKVVPSTYLDTEGAFVTYATRIDRPSIHSOEMLAGOF 654
Oy 661 LOEVSVMAMVK 671
Db 655 LOEVSFMAMVK 665

RESULT 7
Oy 09S66 PRELIMINARY: PRT: 666 AA.
AC 09S66:
Dt 01-MAR-2001 (TREMBLrel. 16, Created)
Dt 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
Dt 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CALPAIN-LIKE PROTEASE 10.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
CX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY;
RX MEDLINE=21359423; PubMed=11375982;
RA Ma H., Fukiage C., Kim Y.H., Duncan M.K., Reed N.A., Shih M.,
RA Azuma M., Shearer T.R.;
RA "Characterization and Expression of Calpain 10. A NOVEL UBIQUITOUS
RT CALPAIN WITH NUCLEAR LOCALIZATION.";
RL J. Biol. Chem. 276:28525-28531(2001).
DR EMBL: AF227909; AACG9736.3; -.
DR HSSP: Q07009; IDP0.
DR MEROPS: C02.018; -.
DR InterPro: IPR001300; Peptidase_C2.
DR InterPro: IPR00169; Thiolprot_act_site.
DR Pfam: PF01067; Calpain_I1; 2.
DR Pfam: PF00648; Peptidase_C2; 1.
DR PRINTS: PR00704; CALPAIN.
DR SMART: SM00230; CysPC; 1.
DR PROSITE: PS00139; THIOL_PROTEASE_CYS; UNKNOWN_1.
KM Protease.
SQ SEQUENCE 666 AA; 74527 MW; A92263A9C8C4947D CRC64;

```

```

Oy 1 MRAGGATPARELFRDAAPADSSLCFCDLSTPLAOFREDITWRPDEICATPRLFPDP 60
Matches 544; Conservative 45; Mismatches 76; Indels 6; Gaps 2;

```

```

Db 1 MRVRAETARELFRDAAPADSSLCFCDLSTPLAOFREDITWRPDEICATPRLFPDP 60
Oy 61 REGOVNGLGLDGMFLCACALOKSRHLLDOVLPQPSMADQEVYRGSTFCRTMDFRNV 120
Db 61 WEGOVNGLGLDGMFLCACALOKSRHLLDOVLPQPSMADQEVYRGSTFCRTMDFRNV 120
Oy 121 EYTTDRPLCLGRCLCFSCQREDVYVPLLEKVVYAKVIGSTEHLMAGVADALVDYTG 180
Db 121 EYTTDRPLCLGRCLCFSCQREDVYVPLLEKVVYAKVIGSTEHLMAGVADALVDYTG 180
Oy 181 LAERNMLKAVAGSGQODPGRGHEHRTQRLHLKQDCLSCVSPFRAGRELGEFHF 240
Db 181 LAERNMLKAVAGSGQODPGRGHEHRTQRLHLKQDCLSCVSPFRAGRELGEFHF 240
Oy 241 IYSDRELQGAQOCILLRLIONPWGRRCMOGLMRBGSGMNOVDAVASSELLQOEGE 300
Db 241 IYSDRELQGAQOCILLRLIONPWGRRCMOGLMRBGSGMNOVDAVASSELLQOEGE 300
Oy 301 FWESEEFRLREDELTVGYPVTEAGHLSLYTERLLCHTRALPGANVKQSGAGCCRNNSG 360
Db 301 FWESEEFRLREDELTVGYPVTEAGHLSLYTERLLCHTRALPGANVKQSGAGCCRNNSG 360
Oy 361 FPNKPFMLRLPESEVCAVYQRRRL---VGTALAG---ASPAVNLPGKYOAV 420
Db 361 FPNKPFMLRLPESEVCAVYQRRRL---VGTALAG---ASPAVNLPGKYOAV 414
Oy 421 GLHMKVEKRRVNLPRVLSMPVAGTACHAYDREYHLRCELSPGYLLAVSTFLKDAPGE 480
Db 421 GLHMKVEKRRVNLPRVLSMPVAGTACHAYDREYHLRCELSPGYLLAVSTFLKDAPGE 474
Oy 481 FLIRVSTGRVSLSAIRAVAKNTPGALPAEMGTVOGLRGSRVQGTAGSRNASTYPT 540
Db 475 FLIRVSTGRVSLSAIRAVAKNTPGALPAEMGTVOGLRGSRVQGTAGSRNASTYPT 534
Oy 541 NRCPEFVPEGPGRCVTRITLHQCPSDTEFFHPIGFHIFQYPEGGRSODAPILLQEBL 600
Db 535 NRCPEFVPEGPGRCVTRITLHQCPSDTEFFHPIGFHIFQYPEGGRSODAPILLQEBL 594
Oy 601 LSCVPHRYAEOEVSRLCLLPAGTYKVVPSTYLDTEGAFVTYATRIDRPSIHSOEMLAGOF 660
Db 595 LSCVPHRYAEOEVSRLCLLPAGTYKVVPSTYLDTEGAFVTYATRIDRPSIHSOEMLAGOF 654
Oy 661 LOEVSVMAMVK 671
Db 655 LOEVSFMAMVK 665

RESULT 8
Oy 09HC94 PRELIMINARY: PRT: 517 AA.
AC 09HC94:
Dt 01-MAR-2001 (TREMBLrel. 16, Created)
Dt 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
Dt 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CALPAIN-LIKE PROTEASE CAPN10C.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=20472315; PubMed=11017071;
RA Horikawa Y., Oda N., Cox N.J., Li X., Orho-Melander M., Hara M.,
RA Hinokio Y., Lindner T.H., Mashima H., Schwarz P.E.,
RA del Bosque-Plata L., Horikawa Y., Oda Y., Yoshituchi I., Collilla S.,
RA Polonsky K.S., Wei S., Concannon P., Iwasaki N., Schnitz J.,
RA Balier L.J., Bogardus C., Groop L., Boerwinkle E., Hanis C.L.,
RA Bell G.I.;
RA "Genetic variation in the gene encoding calpain-10 is associated with
RT type 2 diabetes mellitus.";
RL Nat. Genet. 26:163-175(2000).
DR EMBL: AF089091; AAC17969.1; -.
DR HSSP: Q07009; IDP0.

```

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20472315; PubMed=11017071;
 RA Horikawa Y., Oda N., Cox N.J., Li X., Orho-Melander M., Hara M.,
 RA Hinokio Y., Lindner T.H., Washima H., Schwarz P.E.,
 RA del Bosque-Plata L., Horikawa Y., Oda Y., Yoshuchi I., Colilla S.,
 RA Polonsky K.S., Wei S., Concannon P., Iwasaki N., Schulze J.,
 RA Baier L.J., Bogardus C., Groop L., Boerwinkle E., Hanis C.L.,
 RA Bell G.I.;
 RT "Genetic variation in the gene encoding calpain-10 is associated with
 type 2 diabetes mellitus";
 RL Nat. Genet. 26:163-175(2000).
 DR EMBL; AF089090; AAG17968.1; -;
 DR HSSP; Q07009; 1DFO.
 DR MEROPS; C02.018; -;
 DR InterPro; IPR001300; Peptidase_C2.
 DR InterPro; IPR000169; Thiolprot_act_site.
 DR Pfam; PF01067; Calpain_III; 1.
 DR Pfam; PF00648; Peptidase_C2; 1.
 DR PRINTS; PR00704; CALPAIN.
 DR SMART; SM00230; Cyspc; 1.
 DR PROSITE; PS00139; THIO_L_PROTEASE_CYS; UNKNOWN_1.
 KW Protease.
 SQ SEQUENCE 544 AA; 60656 MW; 17CE7B881A20855E CRC64;

 Query Match 74.0%; Score 2701.5; DB 4; Length 544;
 Best Local Similarity 95.4%; Pred. No. 2.2e-218;
 Matches 498; Conservative 5; Mismatches 6; Indels 13; Gaps

 QY 1 MRAGRGATPARELFRDAAPFAADSSLCFLDSTPLAQFREDITWRPQECATPRLFPDDP 60
 Db 1 MRAGRGATPARELFRDAAPFAADSSLCFLDSTPLAQFREDITWRPQECATPRLFPDDP 60
 QY 61 REGQVKGGLGDCWFLCACAAALQKSRHLLDQVIPPQPSWADQYRGSTFCRIWQGRWV 120
 Db 61 REGQVKGGLGDCWFLCACAAALQKSRHLLDQVIPPQPSWADQYRGSTFCRIWQGRWV 120
 QY 121 EYTTDDRLPCLAGRLCFRCQREDVFWLPLEKVKYAKVHGSYEHLLWAGQVADALVDLTGG 180
 Db 121 EYTTDDRLPCLAGRLCFRCQREDVFWLPLEKVKYAKVHGSYEHLLWAGQVADALVDLTGG 180
 QY 181 LAERNLKGAVSGGQDDPRGRWEHRTCRQLLHKDQCLISCCVLSPRAGARELGEFHAF 240
 Db 181 LAERNLKGAVSGGQDDPRGRWEHRTCRQLLHKDQCLISCCVLSPRAGARELGEFHAF 240
 QY 241 IVSDRLRQAGQGCILLRIQNPGRRCWQGLWREGGEGSWQDAAVASELLSQLEGE 300
 Db 241 IVSDRLRQAGQGCILLRIQNPGRRCWQGLWREGGEGSWQDAAVASELLSQLEGE 300
 QY 301 FWVEEEFLREFDELTVGVPVTEAGHLQSLYTERLLCHTRALPGAWKVGQSAGGCRNNSG 360
 Db 301 FWVEEEFLREFDELTVGVPVTEAGHLQSLYTERLLCHTRALPGAWKVGQSAGGCRNNSG 360
 QY 361 FPSNPKFWLRVSEPEVYIAVLQSRHLHAADWAGARALVGDSTHSWSPASIPGKHQAV 420
 Db 361 FPSNPKFWLRVSEPEVYIAVLQSRHLHAADWAGARALVGDSTHSWSPASIPGKHQAV 420
 QY 421 GLHLWKVKRRVNLPRVLSMPVPVAGTACHAYDREVHLRCELSPGYLLAVPSTFLKDAPGE 480
 Db 421 GLHLWKVKRRVNLPRVLSMPVPVAGTACHAYDREVHLRCELSPGYLLAVPSTFLKDAPGE 480
 QY 481 FLRLVSTGRVSLSAIRAVAK-----NTTFGAAL 509
 Db 481 FLRLVSTGRVSLRALPAASASLCSICSTAGPVPSTPSASI 522

 RESULT 10
 Q9HC93
 ID Q9HC93 PRELIMINARY; PRT: 511 AA.

AC 09HC93;
 DT 01-MAR-2001 (TReMBLrel. 16, Created)
 DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
 DE 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 DE CALPAIN-LIKE PROTEASE CAPN10D.
 OS Homo sapiens (human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NCBI_TaxID=9606;
 RX MEDLINE=20472315; PubMed=11017071;
 RA Horikawa Y., Oda N., Cox N.J., Li X., Orho-Melander M., Hara M.,
 RA Hirokio Y., Lindner T.H., Mashima H., Schwarz P.E.,
 RA del Bosque-Plata L., Horikawa Y., Oda Y., Yoshinuchi I., Collilla S.,
 RA Polonsky K.S., Wei S., Concannon P., Iwasaki N., Schultze J.,
 RA Baier L.J., Bogardus C., Groop L., Boerwinkle E., Hanis C.L.,
 RA Bell G.I.;
 RT "Genetic variation in the gene encoding calpain-10 is associated with
 RT type 2 diabetes mellitus.";
 RL Nat. Genet. 26:163-175(2000).
 DR EMBL: AF089092; AAG17970.1;
 DR HSSP: 007009; 1DFO.
 DR MEROPS: C02.018;
 DR Interpro: IPR001300; Peptidase_C2.
 DR Interpro: IPR000169; Thiolprot_act_site.
 DR Pfam: PF01067; Calpain_III; 1.
 DR Pfam: PF00648; Peptidase_C2; 1.
 DR PRINTS: PR00704; CALPAIN.
 DR SMART: SM00230; Cyspec; 1.
 DR PROSITE: PS00139; THIOI_PROTEASE_CYS; UNKNOWN_1.
 DR Protease.
 KW PROTEASE.
 SQ SEQUENCE 513 AA; 57815 MW; C66DC853F87AEC9C CRC64;

Query Match 74.0%; Score 2699; DB 4; Length 513;
 Best Local Similarity 98.2%; Pred. No. 3.2e-218;
 Matches 496; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 1 MRGRGATPARELFRAAPADSSLEFCDSTPLAQFREDITRRRQETCATRLPPDDP 60
 DB 1 MRGRGATPARELFRAAPADSSLEFCDSTPLAQFREDITRRRQETCATRLPPDDP 60
 QY 61 REGQVQGLLGDGDFLACALOKSRHLDQVIPPQSPVADQEVGSEFTCRIMQGRWY 120
 DB 61 REGQVQGLLGDGDFLACALOKSRHLDQVIPPQSPVADQEVGSEFTCRIMQGRWY 120
 QY 121 EYTTDRPLCLAGRLCFSRQREDVFWPLLEKVAKVHGSYEHLMAGVADALVDTGG 180
 DB 121 EYTTDRPLCLAGRLCFSRQREDVFWPLLEKVAKVHGSYEHLMAGVADALVDTGG 180
 QY 121 EYTTDRPLCLAGRLCFSRQREDVFWPLLEKVAKVHGSYEHLMAGVADALVDTGG 180
 DB 121 EYTTDRPLCLAGRLCFSRQREDVFWPLLEKVAKVHGSYEHLMAGVADALVDTGG 180
 QY 181 LAERNMLKGVASGGQDDRGREHNTCQLHLKQCLISCVLSPRAGARELGEFHA 240
 DB 181 LAERNMLKGVASGGQDDRGREHNTCQLHLKQCLISCVLSPRAGARELGEFHA 240
 QY 241 IYSDLELQOGAGOCILLRIQNPGRRCMOGLMRGSGMSOVDAVAASELLSQOEGE 300
 DB 241 IYSDLELQOGAGOCILLRIQNPGRRCMOGLMRGSGMSOVDAVAASELLSQOEGE 300
 QY 301 FVVEEELREDELTVGYPVTEAGHLQSLYTERLCHTRALPGAVKQSGAGCCNNNG 360
 DB 301 FVVEEELREDELTVGYPVTEAGHLQSLYTERLCHTRALPGAVKQSGAGCCNNNG 360
 QY 361 FVVEEELREDELTVGYPVTEAGHLQSLYTERLCHTRALPGAVKQSGAGCCNNNG 360
 DB 361 FVVEEELREDELTVGYPVTEAGHLQSLYTERLCHTRALPGAVKQSGAGCCNNNG 360
 QY 421 GLHLMKVERKRNPLRVLSMPVAGACAHAYDREVLRCELSPGYTLAVPSFLDAPGE 480
 DB 421 GLHLMKVERKRNPLRVLSMPVAGACAHAYDREVLRCELSPGYTLAVPSFLDAPGE 480
 QY 481 FLIRVFSTGRVSLRAVAVAKTTP 505
 DB 481 FLIRVFSTGRVSLRAVAVAKTTP 505

DB 481 FLIRVFSTGRVSLRAVAVAKTTP 505
 RESULT 11
 ID 09WFE0 PRELIMINARY; PRT; 605 AA.
 AC 09WFE0;
 DT 01-NOV-1999 (TReMBLrel. 12, Created)
 DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
 DE 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 DE CALPAIN-LIKE PROTEASE.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RX MEDLINE=10090;
 RA Braun C., Seifert M., Engel M., Welter C.;
 RA "Identification of a new calpain-like cDNA in mouse lung.";
 RL Submitted (FEB-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF126867; AAD41779.1;
 DR HSSP: 007009; 1DFO.
 DR MEROPS: C02.018;
 DR MGD: MGI:1344392; Capn10.
 DR Interpro: IPR001300; Peptidase_C2.
 DR Interpro: IPR000169; Thiolprot_act_site.
 DR Pfam: PF01067; Calpain_III; 1.
 DR Pfam: PF00648; Peptidase_C2; 1.
 DR PRINTS: PR00704; CALPAIN.
 DR SMART: SM00230; Cyspec; 1.
 DR PROSITE: PS00139; THIOI_PROTEASE_CYS; UNKNOWN_1.
 DR Protease.
 KW PROTEASE.
 SQ SEQUENCE 605 AA; 67788 MW; 85ACEBA43FA0EB12 CRC64;

Query Match 73.7%; Score 2691; DB 11; Length 605;
 Best Local Similarity 80.7%; Pred. No. 1.9e-217;
 Matches 490; Conservative 43; Mismatches 68; Indels 6; Gaps 2;

QY 1 MRGRGATPARELFRAAPADSSLEFCDSTPLAQFREDITRRRQETCATRLPPDDP 60
 DB 1 MRGRGATPARELFRAAPADSSLEFCDSTPLAQFREDITRRRQETCATRLPPDDP 60
 QY 61 REGQVQGLLGDGDFLACALOKSRHLDQVIPPQSPVADQEVGSEFTCRIMQGRWY 120
 DB 61 REGQVQGLLGDGDFLACALOKSRHLDQVIPPQSPVADQEVGSEFTCRIMQGRWY 120
 QY 121 EYTTDRPLCLAGRLCFSRQREDVFWPLLEKVAKVHGSYEHLMAGVADALVDTGG 180
 DB 121 EYTTDRPLCLAGRLCFSRQREDVFWPLLEKVAKVHGSYEHLMAGVADALVDTGG 180
 QY 121 EYTTDRPLCLAGRLCFSRQREDVFWPLLEKVAKVHGSYEHLMAGVADALVDTGG 180
 DB 121 EYTTDRPLCLAGRLCFSRQREDVFWPLLEKVAKVHGSYEHLMAGVADALVDTGG 180
 QY 181 LAERNMLKGVASGGQDDRGREHNTCQLHLKQCLISCVLSPRAGARELGEFHA 240
 DB 181 LAERNMLKGVASGGQDDRGREHNTCQLHLKQCLISCVLSPRAGARELGEFHA 240
 QY 241 IYSDLELQOGAGOCILLRIQNPGRRCMOGLMRGSGMSOVDAVAASELLSQOEGE 300
 DB 241 IYSDLELQOGAGOCILLRIQNPGRRCMOGLMRGSGMSOVDAVAASELLSQOEGE 300
 QY 301 FVVEEELREDELTVGYPVTEAGHLQSLYTERLCHTRALPGAVKQSGAGCCNNNG 360
 DB 301 FVVEEELREDELTVGYPVTEAGHLQSLYTERLCHTRALPGAVKQSGAGCCNNNG 360
 QY 361 FVVEEELREDELTVGYPVTEAGHLQSLYTERLCHTRALPGAVKQSGAGCCNNNG 360
 DB 361 FVVEEELREDELTVGYPVTEAGHLQSLYTERLCHTRALPGAVKQSGAGCCNNNG 360
 QY 421 GLHLMKVERKRNPLRVLSMPVAGACAHAYDREVLRCELSPGYTLAVPSFLDAPGE 480
 DB 421 GLHLMKVERKRNPLRVLSMPVAGACAHAYDREVLRCELSPGYTLAVPSFLDAPGE 480
 QY 481 FLIRVFSTGRVSLRAVAVAKTTP 505
 DB 481 FLIRVFSTGRVSLRAVAVAKTTP 505

QY 481 FLRVSTGRVSLSATRAVAKNTTGAALPAGEMGTGVLGRGWRVGTAGGSRNFASYPT 540
DB 475 FLRVSTGRVSLSATRAVAKNTTGAALPAGEMGTGVLGRGWRVGTAGGSRNFASYPT 534
QY 541 NPFPPSPVGGPCRCVRLTHOHCPCSPDETHPGFHLFQVPEGRCSDAPPLILQEP 600
DB 535 NCLPSPVGGPCRCVRLTHOHCPCSPDETHPGFHLFQVPEGRCSDAPPLILQEP 594
QY 601 LSCVPHR 607
DB 595 LSCVPHR 601
RESULT 12
QYHC92
ID QYHC92 PRELIMINARY; PRT: 444 AA.
AC QYHC92
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE CALPAIN-LIKE PROTEASE CAPN10E.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE=20472315; PubMed=11017071;
RA Horikawa Y., Oda N.J., Li X., Orho-Melander M., Hara M.,
RA Hinokio Y., Lindner T.H., Mashima H., Schwarz P.E.,
RA del Bosque-Plata L., Horikawa Y., Oda Y., Toshiuchi I., Collila S.,
RA Polonsky K.S., Wei S., Concannon P., Iwasaki N., Schulze J.,
RA Baier L.J., Bogardus C., Groop L., Boerwinkle E., Hanis C.L.,
RA Bell G.I.;
RT "Genetic variation in the gene encoding calpain-10 is associated with
RT type 2 diabetes mellitus".
RL Nat. Genet. 26:163-175(2000).
DR EMBL: AF089093; AAG17971.1; -
DR HSSP: Q07009; IDP0.
DR MEROPS: C02.018; -
DR InterPro: IPR001300; Peptidase_C2.
DR InterPro: IPR001169; Thiolprot_act_site.
DR Pfam: PF00648; Peptidase_C2; 1.
DR PRINTS: PR00704; CALPAIN.
DR SMART: SM00230; Cyspc; 1.
DR PROSITE: PS00139; THIOI_PROTEASE_CYS; UNKNOWN_1.
KW Protease.
SQ SEQUENCE 444 AA; 49992 MW; 09527DDOCB06D586 CRC64;
Query Match 64.0%; Score 2336; DB 4; Length 444;
Best Local Similarity 99.8%; Pred. No. 8e-188;
Matches 425; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 MRAGRGATPARELFDAAPFAADSSLCFCDLSTPLAQFREDITWRPQICATPRFLPDDP 60
DB 1 MRAGRGATPARELFDAAPFAADSSLCFCDLSTPLAQFREDITWRPQICATPRFLPDDP 60
QY 61 REGQVKGILLGDCWFLCACAALQKSRHLLDQVIPPQPSWADQEVGSGFTCRWQGRW 120
DB 61 REGQVKGILLGDCWFLCACAALQKSRHLLDQVIPPQPSWADQEVGSGFTCRWQGRW 120
QY 121 EVTTDDRLPCLAGRLCFRSRCQREDVFWLPLEKRYAKVHGYEHLWAGQVADALVLTGG 180
DB 121 EVTTDDRLPCLAGRLCFRSRCQREDVFWLPLEKRYAKVHGYEHLWAGQVADALVLTGG 180
QY 181 LAERNLKGAVSGGQDPRGWEHRTCRQLHLKDKQCLISCCVLSPRAGRELGEHAF 240
DB 181 LAERNLKGAVSGGQDPRGWEHRTCRQLHLKDKQCLISCCVLSPRAGRELGEHAF 240
QY 241 IVSOLRELQOGAGCILLRLTONPWGRRCWGLAREGEGWSDAASVALLSQLEGE 300
DB 241 IVSOLRELQOGAGCILLRLTONPWGRRCWGLAREGEGWSDAASVALLSQLEGE 300

QY 301 FWEEEFELREDELTVGYPVTEAGHLQSLYTERLCHTRALPGANVKGSGAGGCRNNG 360
DB 301 FWEEEFELREDELTVGYPVTEAGHLQSLYTERLCHTRALPGANVKGSGAGGCRNNG 360
QY 361 FFSNPKFWLRSEPVSEVYIAVLQSRSLHAADWAGRALVGDSTSWSPASIPGKHQAV 420
DB 361 FFSNPKFWLRSEPVSEVYIAVLQSRSLHAADWAGRALVGDSTSWSPASIPGKHQAV 420
QY 421 GLHLWK 426
DB 421 GLHLWK 426
RESULT 13
QYCPY2
ID QYCPY2 PRELIMINARY; PRT: 264 AA.
AC QYCPY2
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE CALPAIN 10.
GN CAPN10.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Glissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baidarelli R., Barsh G.,
RA Blake J., Boiffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Marzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wyushaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AK013497; BAB28883.1; -
DR EMBL: AK002548; BAB22179.1; -
DR HSSP: Q07009; IDP0.
DR MEROPS: C02.018; -
DR MGD: MGI:1344392; Capn10.
DR InterPro: IPR001300; Peptidase_C2.
DR InterPro: IPR001169; Thiolprot_act_site.
DR Pfam: PF00648; Peptidase_C2; 1.
DR PRINTS: PR00704; CALPAIN.
DR SMART: SM00230; Cyspc; 1.
DR PROSITE: PS00139; THIOI_PROTEASE_CYS; UNKNOWN_1.
SQ SEQUENCE 264 AA; 29946 MW; EBA53040A47E65F2 CRC64;
Query Match 29.9%; Score 1091; DB 11; Length 264;
Best Local Similarity 86.1%; Pred. No. 1.4e-83;
Matches 198; Conservative 11; Mismatches 21; Indels 0; Gaps 0;
QY 1 MRAGRGATPARELFDAAPFAADSSLCFCDLSTPLAQFREDITWRPQICATPRFLPDDP 60
DB 1 MRAGRGATPARELFDAAPFAADSSLCFCDLSTPLAQFREDITWRPQICATPRFLPDDP 60

Page 9

03ER56: 01-MAR-2001 (TREMblrel. 16, Created)
 DT 01-MAR-2001 (TREMblrel. 16, last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, last annotation update)
 DE CALPAIN 12.
 GN CAPN12.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-20422668; PubMed-10964513;
 RA Dear T.N., Meier N.T., Boehm T.;
 RT "Gene structure, chromosomal localization and expression pattern of
 RT Capn12, a new member of the calpain large subunit gene family.";
 RL Genomics 68:152-160(2000).
 DR EMBL; AJ289241; CAC10066.1; -.
 DR HSSP; Q07009; IDFO.
 DR MEROPS: C02.017; -.
 DR MGD; MGI:1891369; Capn12.
 DR InterPro: IPR002048; EF-hand.
 DR InterPro: IPR001300; Peptidase_C2.
 DR Pfam; PF01067; Calpain_III; 1.
 DR Pfam; PF00648; Peptidase_C2; 1.
 DR PRINTS; PR00704; CALPAIN.
 DR SMART; SM00230; Cyspc; 1.
 DR PROSITE; PS00018; EF_HAND; UNKNOWN.1.
 DR SEQUENCE 720 AA; 8038 MW; 37C07BDF0145E531 CRC64;

Query Match	18.0%	Score 657.5;	DB 11;	Length 720;
Best Local Similarity	32.1%	Pred. No. 1.3e-46;		
Matches 170;	Conservative 71;	Mismatches 183;	Indels 105;	Gaps 14

QY	13	LEFRAAEPAADSLFCEDLSPELAEQFRDITWRROEICATPRLEPPDDPREQVOKGILGD	72
Db	45	LEFRDPCFAPGDALGDKLGPDSSEKAKGVEMKRNEHCAPERQFICEDMSRFDVCGGSLGN	104
QY	73	CMFLCACAAALOKSHHLDQVTPRGQPSMAODEYUGSTCRIMQGRNVEVTTDRPLCLA	133
Db	105	CMILAAAASLTLRYLLRYRVPQO--GFODGYGVHFQIMQFGRWADVVDDDKLRYRE	162
QY	133	GRLECFSCROREDVYMLPTEKVTYAKVGSYHEULNAGOVADLYDTGLGLERNMLKCVAG	199
Db	163	GKLMFVSEBQNEFMETAELEKAYAKKLGSTVEVMGSHNEFMVDPFGVGVLYLR----	218
QY	193	SGGQODRPRGWEHNTCRQLH-LKDOCLISCVYLSPPRAGAE--LGEFNATYSDELREL	248
Db	219	-----QMTPG-----VFALRHRLAKESLYGATLSDRGEITDEGLYKGNHISTYGNHKM	265
QY	249	QGAAGOCILILLRIONPFGRCWQSLMRBEGGWSQVDAVAASELLSOLQDEGEFVVEEER	306
Db	270	-SIGFTYRLRLRLNPMGGRVEMSGPMWDCSPRMIMLSEWMDALYVRKEGEFEMELQDF	328
QY	309	LRFEDELTVGPRVTEAGHLSLYTERL-----LCPTRLRPGMWYVGSAGCCRRNS-G	366
Db	329	LTHERNTV-----QICSLSEFVLGAPSPBAGGWHIHLPGGRWVGNFSGSGSPAEEN	378
QY	361	FPSNPKEMLRYSESE-----YIAVLQSRSLH	388
Db	379	FMTNPGQRLTLLEPDEDEDDODEGPRGWSGAGARPARPAGGRVYKCTVLLSLQRRR-	437
QY	389	AADMAGARALVGDSTSWSPASTPGKNYAVGLH-----LWKVKRRVRYNPRVL	438
Db	438	-----CLRAKGLT-----YLTGFGHVQIPEELLDLWDSPRSALLPGLL	477
QY	439	SMRPVATAGHAUDREYHLRCESLPGYUAVLPSFLFLDADGCEFLLRPTS	487
Db	478	R-----ADRSVCARADVSRRCRLRPGHYUAVLWVSSARVGDADFLRLTFS	522

Thu Sep 19, 15:10:44 2002

us-09-768-877-2.sepi9.rsp

Page 1

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 19, 2002, 14:50:01 ; Search time 13.56 seconds

(without alignments)
1918.845 Million cell updates/sec

Title: US-09-768-877-2

Sequence: 1 MRAGRGATPARELFRDAAP.....SOEMLGQFLQEVSWAVAKT 672

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched:

105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters:

105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08

Maximum Match 1008
Listing first 45 summaries

Database : SWISSPROT_40.*

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	644.5	17.7	702	1	CANB_HUMAN
2	642	17.6	705	1	CANX_CHICK
3	631	17.3	714	1	CANL_PIG
4	629	17.2	700	1	CAN2_MOUSE
5	621.5	17.0	713	1	CAN1_RAT
6	618.5	16.9	714	1	CAN1_HUMAN
7	618	16.9	821	1	CAN3_HUMAN
8	615.5	16.9	713	1	CAN1_MOUSE
9	614	16.8	700	1	CAN2_RAT
10	605.5	16.6	821	1	CAN3_MOUSE
11	604	16.6	700	1	CAN2_HUMAN
12	599.5	16.4	821	1	CAN3_RAT
13	597.5	16.4	700	1	CAN2_CHICK
14	589.5	16.2	810	1	CAN3_CHICK
15	586	16.1	828	1	CAN1_MOUSE
16	579	15.9	653	1	CAN1_MOUSE
17	545	14.9	758	1	CAN1_MOUSE
18	512.5	14.0	641	1	CAN6_HUMAN
19	483.5	13.3	641	1	CAN6_MOUSE
20	376.5	10.3	1597	1	SOL_DROME
21	268.5	7.4	422	1	CAN2_MOUSE
22	266.5	7.3	374	1	CAN3_PIG
23	114	3.1	528	1	MDLC_PSEPU
24	111	3.0	875	1	NETR_HUMAN
25	110	3.0	3519	1	OL56_STRAT
26	109.5	3.0	116	1	CAN3_BOVIN
27	106	2.9	324	1	CAN2_PIG
28	106	2.9	761	1	NETR_MOUSE
29	105.5	2.9	687	1	TGIC_BOVIN
30	104.5	2.9	673	1	Y552_HUMAN
31	102.5	2.8	686	1	TGIC_MOUSE
32	102.5	2.8	687	1	TGIC_MOUSE
33	102	2.8	573	1	AMH2_HUMAN

ALIGNMENTS

RESULT	1	STANDARD	PRT	702 AA.
ID	CANB_HUMAN			
AC	09UMG6;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Calpain II (EC 3.4.22.17) (Calcium-activated neutral proteinase II)			
DE	(CANP II).			
GN	CANP11.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=9333989; PubMed=10409436;			
RA	Dear T.N., Moller A., Boehm T.;			
RT	"CAPN11: A calpain with high mRNA levels in testis and located on			
RT	chromosome 6."			
RT	Genomics 59:243-247(1999).			
CC	- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which			
CC	catalyze limited proteolysis of substrates involved in			
CC	cytoskeletal remodelling and signal transduction.			
CC	- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-I-Xaa, Met-I-Xaa or			
CC	Arg-I-Xaa with leu or val as the P2 residue.			
CC	- SUBUNIT: Heterodimer of a large (catalytic) and a small			
CC	(regulatory) subunit.			
CC	- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN TESTIS.			
CC	- SIMILARITY: CONTAINS 4 EF-HAND CALCIUM-BINDING DOMAINS.			
CC	- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C2; ALSO KNOWN AS THE			
CC	CALPAIN FAMILY OF THIOL PROTEASES.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@sib-sib.ch).			
CC	-----			
DR	EMBL: AJ242832; CAB52137.1; -			
DR	HSSP: P17655; 10KV.			
DR	MIM: 604822; -			
DR	InterPro: IPR002048; EF-hand.			
DR	InterPro: IPR001300; Peptidase_C2.			
DR	InterPro: IPR0010169; Thiolprot_act_site.			
DR	Pfam: PF01067; Calpain_III; 1.			
DR	Pfam: PF00036; ehand; 2.			
DR	Pfam: PF00648; Peptidase_C2; 1.			
DR	PRINTS: PR00704; CALPAIN.			
DR	SMART: SM00230; Cyasp; 1.			
DR	SMART: SM00054; Efn; 2.			
DR	PROSITE: PS00018; EF_HAND; UNKNOWN 2.			
DR	PROSITE: PS00139; THIOL_PROTEASE_CYS; 1.			
DR	PROSITE: PS00639; THIOL_PROTEASE_HIS; FALSE_NEG.			

34	102	2.8	2504	1	FAS_HUMAN	P49327	homo sapien
35	101.5	2.8	302	1	CAN1_RABIT	P06815	Oryctolagus
36	100.5	2.8	1463	1	PAZR_BOVIN	P49259	bos taurus
37	99	2.7	517	1	GLPK_MYCTU	O69664	mycobacteri
38	96.5	2.6	429	1	KNIR_DROME	P10734	drosophila
39	95.5	2.6	4303	1	PKD1_HUMAN	P98161	homo sapien
40	95	2.6	457	1	TMS5_HUMAN	O9H353	homo sapien
41	94.5	2.6	808	1	FGFR_MOUSE	O03142	mus sapien
42	94.5	2.6	941	1	GCSP_MYCTU	O50601	mycobacteri
43	94	2.6	748	1	PTIP_SALTY	P37178	salmonella
44	93.5	2.6	689	1	TGIC_CAVCU	P08587	cavia cutle
45	93	2.5	207	1	CAN2_BOVIN	O27971	bos taurus

DR PROSITE; PS00640; THIOLEPROTEASE ASN; FALSE_NEG.
 KW Hydrolyase; Thiol protease; Calcium-binding; Multigene family.
 FT DOMAIN 17 207 THIOLEPROTEASE DOMAIN I.
 FT DOMAIN 208 352 THIOLEPROTEASE DOMAIN II.
 FT DOMAIN 353 514 DOMAIN III, C2-LIKE DOMAIN.
 FT DOMAIN 515 530 LINKER.
 FT DOMAIN 531 701 DOMAIN IV.
 FT CA_BIND 586 597 EF-HAND 1.
 FT CA_BIND 616 627 EF-HAND 2.
 FT DOMAIN 651 662 ANCESTRAL CALCIUM SITE 3 (POTENTIAL).
 FT DOMAIN 681 692 ANCESTRAL CALCIUM SITE 4 (POTENTIAL).
 FT ACT_SITE 102 102 BY SIMILARITY.
 FT ACT_SITE 259 259 BY SIMILARITY.
 FT ACT_SITE 283 283 BY SIMILARITY.
 SQ SEQUENCE 702 AA; 80582 MW; 3A618682B9674FAE CRC64;

Query Match 17.7%; Score 644.5; DB 1; Length 702;
 Best Local Similarity 33.7%; Pred. No. 1.6e-43;
 Matches 168; Conservative 71; Mismatches 201; Indels 59; Gaps 15;

QY 12 ELFRDAAPPAASL-FCDLSTPLAFREDITWRRPQICATPRLPDDPREQGVKGL 70
 DB 41 ELFEDPLPAPPSGLGKDLG-PNSKNVQNSWOPADINPLFMDGSPDTCOGIL 99
 QY 71 GDCWFLCACAALOKSRHLLDQVTPGQPSWADQYRGSEFCRIWQFGRVVEVTDRLPC 130
 DB 100 GDCWLLAAIGSLTCTPCLLYRVVPRGOS--FKKNYAGIPHFQIQWQGVNVVVDRLPT 157
 QY 131 LAGRLCFSCQREDVFWLPLEKLYAKVHGSYELHLAGQVADALVDLTGGLAERWNLKV 190
 DB 158 KNDKLVFVHSTSEFWSALLERKAYAKLSGYEALSGSGSTMEGLDFTGGVAQSFQLO-- 215
 QY 191 AGSGGQDRPRGWEHRTCRQLLKLQCLISCCVLSPRAGARE-----LGEFHAFIVSD 244
 DB 216 -----RPPNLLRLRKA--VERSLMGCSLEVTSELESMTDKMLVRGHAYSVTG 265
 QY 245 LRELOQAGOCILLRIQNPWRRCQGLWREGGEGNSQVDAVASSELLSQLEGFEWVE 304
 DB 266 LDVH-YRGKMETLIRVNPWRIENGAWGASAREWEVVASDIOMQLLHKTEGFWMS 324
 QY 305 EEFLEFEDELTVG--YPTVEAGHLQSLTERLLCHTRALPGWVWGSGAGGCRNNSG-F 361
 DB 325 YQDFLNNFILLEICNTPTDLSGDKSYW-----HTTYESWRGSGAGGCRNHPGTF 378
 QY 362 PSNPKFWLAVSE---PSEYVIAVLRSLRLHAADWAGR---ARALVGDSTHSWSPASIPCK 415
 DB 379 WTNQPKISLPGDDPED-----DAEGNVVVCTCLVALMKNHARQQA 424
 QY 416 HYQAVGLHLWKVKRNVNLPV-LSMPVYAGTACHAY-----DREVHLRCELSPGYLA 468
 DB 425 QLTGICFVLYAVPKFQNIQDVLHKKFTKYQDHGFSEIFTNREVSSQLRPPGEYII 484
 QY 469 VPSTFLKDPAGFEFLRVFS 487
 DB 485 IPSTFPHRDAFLRVFT 503

RESULT 2
 ID CANX_CHICK STANDARD; PRT; 705 AA.
 AC P00789;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Calpain, large [catalytic] subunit (EC 3.4.22.17) (Calcium-activated neutral proteinase) (CAMP) (Mu/M-type).
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE=85061606; PubMed=6095110;
 RA Ohno S., Emori Y., Imajoh S., Kawasaki H., Kisaragi M., Suzuki K.;
 RT "Evolutionary origin of a calcium-dependent protease by fusion of
 RL genes for a thiol protease and a calcium-binding protein?";
 RN Nature 312:566-570(1984).
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86082358; PubMed=3000828;
 RA Emori Y., Ohno S., Tobita M., Suzuki K.;
 RT "Gene structure of calcium-dependent protease retains the ancestral
 RL organization of the calcium-binding protein gene.";
 RN FEBS Lett. 194:249-252(1986).
 RP CALCIUM-BINDING DATA.
 RX MEDLINE=87279982; PubMed=3038855;
 RA Minami Y., Emori Y., Kawasaki H., Suzuki K.;
 RT "E-F hand structure-domain of calcium-activated neutral protease
 RL (CAMP) can bind Ca2+ ions.";
 RN J. Biochem. 101:889-895(1987).
 RP CHARACTERIZATION.
 RX MEDLINE=95260862; PubMed=7742367;
 RA Sorimachi H., Tsukahara T., Okada-Ban M., Sugita H., Ishiura S.,
 RA Suzuki K.;
 RT "Identification of a third ubiquitous calpain species -- chicken
 RL muscle expresses four distinct calpains.";
 CC Biochim. Biophys. Acta 1261:381-393(1995).
 CC -!- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which
 CC catalyze limited proteolysis of substrates involved in
 CC cytoskeletal remodelling and signal transduction.
 CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-I-Xaa, Met-I-Xaa or
 CC Arg-I-Xaa with Leu or Val as the P2 residue.
 CC -!- COFACTOR: Binds 3 calcium ions.
 CC -!- ENZYME REGULATION: Activated by micromolar concentrations of
 CC calcium and inhibited by calpastatin.
 CC -!- SUBUNIT: Heterodimer of large (catalytic) and a small (regulatory)
 CC subunit.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma
 CC membrane upon Ca++ binding (By similarity).
 CC -!- TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED.
 CC -!- PTM: THE N-TERMINUS IS BLOCKED.
 CC -!- SIMILARITY: Contains 5 EF-hand calcium-binding domains.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C2; ALSO KNOWN AS THE
 CC CALPAIN FAMILY OF THIOLEPROTEASES.
 CC -!- CAUTION: THIS PROTEIN WAS PREVIOUSLY THOUGHT TO BE M-CALPAIN BUT
 CC HAS SINCE BEEN FOUND TO BE AN INTERMEDIATE FORM BETWEEN THE M AND
 CC MU TYPES.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; X01415; CAA25658.1; -
 PIR; A00979; CICHV.
 HSP; P17655; IDKV.
 MEROPS; C02.003; -
 InterPro; IPR002048; EF-hand.
 InterPro; IPR001300; Peptidase_C2.
 InterPro; IPR000169; Thiolproct_act_site.
 Pfam; PF01067; Calpain_III; 1.
 Pfam; PF00036; ehand; 2.
 Pfam; PF00648; Peptidase_C2; 1.
 PRINTS; PR00704; CALPAIN.
 SMART; SM00230; Cyspc; 1.
 PROSITE; PS00018; EF_HAND; 1.
 PROSITE; PS00139; THIOLEPROTEASE_CYS; 1.
 PROSITE; PS00639; THIOLEPROTEASE_HIS; FALSE_NEG.
 PROSITE; PS00640; THIOLEPROTEASE_ASN; FALSE_NEG.

KM Hydrolyase; Thiol protease; Calcium-binding; Multigene family.
 FT DOMAIN 23 213 THIOI PROTEASE DOMAIN I.
 FT DOMAIN 214 358 THIOI PROTEASE DOMAIN II.
 FT DOMAIN 359 517 THIOI PROTEASE DOMAIN III.
 FT DOMAIN 518 533 LINKER.
 FT DOMAIN 534 704 DOMAIN IV.
 FT CA_BIND 545 556 EF-HAND 1.
 FT CA_BIND 589 600 EF-HAND 2.
 FT CA_BIND 619 630 EF-HAND 3.
 FT DOMAIN 654 665 ANCESTRAL CALCIUM SITE 4 (POTENTIAL).
 FT DOMAIN 684 695 ANCESTRAL CALCIUM SITE 5 (POTENTIAL).
 FT ACT_SITE 108 108 BY SIMILARITY.
 FT ACT_SITE 265 265 BY SIMILARITY.
 FT ACT_SITE 289 289 BY SIMILARITY.
 SO SEQUENCE 705 AA; ABCDDC5629848AA CRC64;

Query Match 17.6%; Score 642; DB 1; Length 705;
 Best Local Similarity 32.0%; Pred. No. 2.6e-43;
 Matches 162; Conservative 78; Mismatches 188; Indels 78; Gaps 16;

QY 13 LFRDAPPAADSSL-FCDSLPPLAFREDITRRPOELCATRLPPDDREGQVKGGLG 71
 DB 48 LFRDQFPAGPPLALGFKELG-PYSSKTRGVENKRSSELVDPOFTYVGATRDICGAGLG 106
 QY 72 DCMFLACALAKSRNHLDOVTPGQPSWADQERGSFTCRIMQGRWVEVTTDRLPCL 131
 DB 107 DCMFLAAGSLNLNELLHRVPHGOS--FQDYAGIFHFOIWOGEWVYVDDLLPYK 164
 QY 132 AGRCLFSCQREDVWPLLEKRVAKVHSGVEHLNAGOVADALVDLTGLAEKMLKGYA 191
 DB 165 DGEILFVHSAECTEFWSALLERAKVAKINGCYESLGGSTTEGEEDFTGVAEMTDLK-- 221
 QY 192 GSGGQDRGRNHRHTCQLH--LKDCLISCCVLSPPAGARELGEF--HAF 240
 DB 222 -----RAPRMGHIIKRLALERGLSLCCSIDITSAFDMENATFKIKVGHAY 267
 QY 241 IVSDLREL--QSOAGOCILLRIQPMGRRCMOGLMREGESMSOVDAVASSELLSOLOE 298
 DB 268 SVTAFKQVNYNGQGOQ---LIRIRPMQOVETGMSBSSWMDINDSDRELQIKMED 324
 QY 299 GEFVVEEPELREDELTGVGPVTEAGHLQSLYTERL-LCHTRALPGAMVKGQSGAGCCHN 357
 DB 325 GEFVMSFDFMFEFSLRICNLTPDA-----LTKDELSMHTQVEGRTWRSGTAGGCCHN 379
 QY 358 N-SGSPSPKFMRLVSESE-----YIATVQSRKLAADWAGRALVGDSTHT- 405
 DB 380 NPATFWINPQFKILREDDDDPGDEVAACSLVALMOKHR-----RRERVGGDMHTI 432
 QY 406 SMSPPASIPGKHQAVGLHMK---VEKRRVNLPRVLNMPVAGTACHAYDEVELHRCLE 461
 DB 433 GFAYVEVEEAGSGNVHLKKDFLARNOSKARSEFIML-----REVSNQRL 480
 QY 462 SPGYTLAVPSTFLDAPGEFLLRVFS 487
 DB 481 PPEGIYVPSFTEPHKEADFLIRVET 506

RESULT 3
 CANI_PIG STANDARD: PRT: 714 AA.
 AC P35750; Q29600; Q9NOM6;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Calpain 1, large [catalytic] subunit (BC 3.4.22.17) (Calcium-activated neutral proteinase) (CANP) (Mu-type) (mucAMP) (Microtubular-calpain).
 GN CAPN1.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
 OX NCBI_TaxID=9623;
 RN [1]

RP SEQUENCE FROM N.A.
 RA Smith T.P.L., Simmen F.A., Vallet J.A.;
 RT "Rapid Communication: nucleotide sequences of two isoforms of porcine micromolar calcium-activated neutral proteinase 1 (mucalpain) cDNA.";
 RL Submitted (MAY-2000) to the EMBL/Genbank/DBJ databases.
 [2]
 RP SEQUENCE OF 326-415 FROM N.A.
 RC TISSUE-Small intestine;
 RA Winteroe A.K., Fredholm M., Davies W.;
 RL Submitted (AUG-1995) to the EMBL/Genbank/DBJ databases.
 [3]
 RP SEQUENCE OF 528-623 FROM N.A.
 RX TISSUE-Skeletal muscle;
 RX MEDLINE-94146155; PubMed-8312396;
 RA Sun W., Ji S.O., Ebert P.J., Bidwell C.A., Hancock D.L.;
 RT "Cloning the partial cDNAs of mu-calpain and m-calpain from porcine skeletal muscle.";
 RL Biochimie 75:931-936(1993).
 CC -1- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which catalyze limited proteolysis of substrates involved in cytoskeletal remodeling and signal transduction.
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-I-Xaa, Met-I-Xaa or Arg-I-Xaa with Ieu or Val as the p2 residue.
 CC -1- COFACTOR: Binds 3 calcium ions (By similarity).
 CC -1- ENZYME REGULATION: Activated by micromolar concentrations of calcium and inhibited by calpastatin.
 CC -1- SUBUNIT: Heterodimer of a large (catalytic) and a small (regulatory) subunit.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma membrane upon Ca++ binding (By similarity).
 CC -1- SIMILARITY: Contains 5 EF-hand calcium-binding domains.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C2; ALSO KNOWN AS THE CALPAIN FAMILY OF THIOI PROTEASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AF263610; AAF73444.1;
 DR EMBL: F14611; CA23154.1;
 DR EMBL: 001180; AAG5125.1;
 DR HSSP: P17655; IDKV.
 DR MEROPS: C02.001;
 DR InterPro: IPR002048; EF-hand.
 DR InterPro: IPR001300; Peptidase_C2.
 DR InterPro: IPR00169; Thiolprotease.
 DR Pfam: PF01067; Calpain_I; 1.
 DR Pfam: PF00648; Calpain_III; 1.
 DR Pfam: PF00648; Peptidase_C2; 1.
 DR PRINTS: PR00704; CALPAIN.
 DR SMART: SM00230; Cysrc; 1.
 DR SMART: SM00054; Efb; 2.
 DR PROSITE: PS00016; EF-HAND; 2.
 DR PROSITE: PS00139; THIOI-PROTEASE_CYS; 1.
 DR PROSITE: PS00639; THIOI-PROTEASE_HIS; FALSE NEG.
 DR PROSITE: PS00640; THIOI-PROTEASE_ASN; FALSE NEG.
 KW Hydrolyase; Thiol protease; Calcium-binding; Multigene family.
 FT DOMAIN 221 365 THIOI PROTEASE DOMAIN I.
 FT DOMAIN 366 526 THIOI PROTEASE DOMAIN II.
 FT DOMAIN 527 542 LINKER.
 FT DOMAIN 543 713 DOMAIN IV.
 FT CA_BIND 554 565 EF-HAND 1.
 FT CA_BIND 598 609 EF-HAND 2.
 FT CA_BIND 628 639 EF-HAND 3.
 FT CA_BIND 663 674 ANCESTRAL CALCIUM SITE 4 (POTENTIAL).
 FT DOMAIN 693 704 ANCESTRAL CALCIUM SITE 5 (POTENTIAL).
 FT ACT_SITE 115 115 BY SIMILARITY.
 FT ACT_SITE 272 272 BY SIMILARITY.
 FT ACT_SITE 296 296 BY SIMILARITY.

SPRAIN-BALB/C;
Ozaki Y.;
Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
[3]
SEQUENCE FROM N.A.
TISSUE-CNS;
Glass J.D., Nash N.R., Dry I., Culver D., Wesselingh S.;
Cloning of m-calpain from mouse nervous system.;
Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which
catalyze limited proteolysis of substrates involved in
CC cytoskeletal remodeling and signal transduction (By similarity).
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-|-Xaa, Met-|-Xaa or
CC Arg-|-Xaa with Leu or Val as the P2 residue.
CC -1- COFACTOR: Binds 3 calcium ions.
CC -1- ENZYME REGULATION: Activated by 200-1000 micromolar concentrations
of calcium and inhibited by calpastatin.
CC -1- SUBUNIT: Heterodimer of a large (catalytic) and a small
(regulatory) subunit.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma
membrane upon Ca++ binding.
CC -1- SIMILARITY: CONTAINS 5 EF-HAND CALCIUM-BINDING DOMAINS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C2; ALSO KNOWN AS THE
CALPAIN FAMILY OF THIOL PROTEASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; Y10139; CAA71227.1; -;
DR EMBL; D38117; BAA32964.1; -;
DR EMBL; AF015038; AAB94029.1; -;
DR HSSP; P17655; 1DKV.
DR MEROPS; C02.002; -;
DR MGD; MGI:88264; Capn2.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR001300; Peptidase_C2.
DR InterPro; IPR000169; Thiolprot_act_site.
DR Pfam; PF01067; Calpain_III; 1.
DR Pfam; PF00036; efband; 3.
DR Pfam; PF00648; Peptidase_C2; 1.
DR PRINTS; PR00704; CALPAIN.
DR SMART; SM00230; Cyspc; 1.
DR SMART; SM00054; Efh; 2.
DR PROSITE; PS00018; EF_HAND; 2.
DR PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
DR PROSITE; PS00639; THIOL_PROTEASE_HIS; FALSE_NEG.
DR PROSITE; PS00640; THIOL_PROTEASE_ASN; FALSE_NEG.
KW Hydroxylase; Thiol protease; Calcium-binding; Repeat; Multigene family.
FT PROPEP 1 19
FT CHAIN 20 700
FT DOMAIN 20 210
FT DOMAIN 211 355
FT DOMAIN 356 514
FT DOMAIN 515 529
FT DOMAIN 530 700
FT CA_BIND 541 552
FT CA_BIND 585 596
FT CA_BIND 615 626
FT DOMAIN 650 661
FT DOMAIN 680 691
FT ACT_SITE 105 105
FT ACT_SITE 262 262
FT ACT_SITE 286 286
FT CONFLICT 194 194
FT CONFLICT 212 212
FT CONFLICT 402 402
FT SEQUENCE 700 AA; 79871 MW; 682146B290968316 CRC64;

Query Match 17.2%; Score 629; DB 1; Length 700;
 Best Local Similarity 30.7%; Pred. No. 2.7e-42;
 Matches 158; Conservative 78; Mismatches 207; Indels 72; Gaps 12;

QY 10 ARELEDAAPPAADSLCDLSTPLAQRREDITWRRRPOICATPPLFPDDPREGQKGL 69
 DB 42 AGALFQDPSFPLALPSLCKELGYSSTKRGLEMRPEICADPOFIIGATRTDICOA 101
 QY 70 LGDCWFLCACAALOKSRHLLDQVTPPGQPSWADDERGSCFCRIQWGFVETDDSLP 129
 DB 102 LGDCWFLAALAIASLTINEELIARVPPQS--FQENVAGIFHFQWQGEWEVYDDSLP 159
 QY 130 CLARLFCRQREDVWPLPLEKVVAVHGSYTEHMGVADALVDLTGGLAERWNLKG 189
 DB 160 TKDELLFVHSAEGSEFMSALLEKAYAKINCYEALSGATTGEPEDTGGIAEYELR- 218
 QY 190 VAGSGQODRPRGHEHRTCRQLHLKQOCLSCVLSRAGAR-----LGEHAFIVS 243
 DB 219 -----KRPPLFKTIQKA--LEKSLGCSIDITSADSAVYQKLVKGHAYSVT 267
 QY 244 DLKELQAOACQILLRIQNWGRRCQGLMRGEGMSQVDAVASLSQLGGEFVW 303
 DB 268 GAEEVE--SGSLQKLRIRNPMQGVETGKMNDCPSMNTYDPRANLTERQEDGEFWM 326
 QY 304 EEEFLREFDELTVGYVTEAGHLQSLYTERLCHT-----RALPGAWMGQSGAGCRN 357
 DB 327 SFSDFLHHSRL-----EICNLTPDLICDSYKKMKLTFMGDNMRGSAAGCRN 376
 QY 358 -NSGFSNPFVLRVSESE-----VYIAVLRSLRLHADNAGRALVGDSDHS 406
 DB 377 YPRTFMNPOYLKLEDEDEDEDEGERCTFLVGLQKHR-----RQRKMGEDMHTI 429
 QY 407 W-----SPASIPCKHYQAVGLHLMKVEKRRVNLPRVLSMPVAGTACHADREYHLC 461
 DB 430 GEGIVPEPELTQJTNHLCRNFELFTFRABERSDPTFNL-----REVLNRFKL 477
 QY 462 SPGYLAVSTELKADAPERGLAVFSTGRVLSAT 496
 DB 478 PGEIVLPSTEPHKKDGCILRFSEKRDYQAV 512

RESULT 5
 CAN1_RAT
 ID CAN1_RAT STANDARD; PRT; 713 AA.
 AC P97571;
 DT 16-OCT-2001 (Rel. 40; Created)
 DT 16-OCT-2001 (Rel. 40; Last sequence update)
 DE 16-OCT-2001 (Rel. 40; Last annotation update)
 DE Calpain 1, large [catalytic] subunit (EC 3.4.22.17) (Calcium-activated neutral protease) (CAMP) (Mu-type) (mucAMP) (Micromolar-calpain).
 GN CAMP1 OR CLS1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_Taxid=10116;
 RN
 RP SEQUENCE FROM N.A.
 RA MEDLINE-97107433; PubMed-8950173;
 RA Sotomachi H., Amano S., Ishiura S., Suzuki K.;
 RT "Primary sequences of rat mu-calpain large and small subunits are, respectively, moderately and highly similar to those of human.";
 RL Blochm. Biophys. Acta 1309:37-41(1996).
 CC -1- FUNCTION: Calcium regulated non-lysosomal thiol-protease which catalyze limited proteolysis of substrates involved in cytoskeletal remodeling and signal transduction.
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-I-Xaa, Met-I-Xaa or Arg-I-Xaa with Leu or Val as the p2 residue.
 CC -1- COFACTOR: Binds 3 calcium ions (by similarity).
 CC -1- ENZYME REGULATION: Activated by micromolar concentrations of calcium and inhibited by calpastatin.
 CC -1- SUBUNIT: Heterodimer of a large (catalytic) and a small (regulatory) subunit.

CC -1- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma membrane upon Ca++ binding (By similarity).
 CC -1- SIMILARITY: Contains 5 EF-hand calcium-binding domains.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C2; ALSO KNOWN AS THE CALPAIN FAMILY OF THIOL PROTEASES.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.1sb-sib.ch/announce/> or send an email to license@1sb-sib.ch).
 CC EMBL; U53858; AAC53001.1; -
 CC HSSP; P17655; IDKV.
 DR MEROPS; C02.001; -
 DR Interpro; IPR002048; EF-hand.
 DR Interpro; IPR001300; Peptidase_C2.
 DR Interpro; IPR001693; Thiolprol_act_site.
 DR Pfam; PF01067; Calpain_I; 1.
 DR Pfam; PF00036; ehand; 3.
 DR Pfam; PF00648; Peptidase_C2; 1.
 DR PRINTS; PR00704; CALPAIN.
 DR SMART; SM00230; CysPc; 1.
 DR SMART; SM00054; Efh; 2.
 DR PROSITE; PS00018; EF_HAND; 2.
 DR PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
 DR PROSITE; PS00639; THIOL_PROTEASE_HIS; FALSE NEG.
 DR PROSITE; PS00640; THIOL_PROTEASE_ASP; FALSE NEG.
 KW Calcium-binding; Hydrolase; Thiol protease.
 FT DOMAIN 30 220
 FT 221 365 THIOL PROTEASE DOMAIN I.
 FT DOMAIN 366 525 THIOL PROTEASE DOMAIN II.
 FT DOMAIN 526 541 LINKER.
 FT DOMAIN 542 712 DOMAIN III, C2-LIKE DOMAIN.
 FT CA_BIND 553 564
 FT CA_BIND 564 597
 FT CA_BIND 597 608
 FT CA_BIND 627 638
 FT DOMAIN 662 673
 FT DOMAIN 692 703 ANCESTRAL CALCIUM SITE 3 (POTENTIAL).
 FT ACT_SITE 115 115 ANCESTRAL CALCIUM SITE 4 (POTENTIAL).
 FT ACT_SITE 272 272 BY SIMILARITY.
 FT ACT_SITE 296 296 BY SIMILARITY.
 SQ SEQUENCE 713 AA; 82118 MW; 6E664600B0FABEBB CRC64;

Query Match 17.0%; Score 621.5; DB 1; Length 713;
 Best Local Similarity 31.8%; Pred. No. 1.1e-41;
 Matches 161; Conservative 80; Mismatches 188; Indels 77; Gaps 15;

QY 13 LFRDAAPPAADSL-FCDSLTPLAQFREDITWRRRPOICATPPLFPDDPREGQKGLG 71
 DB 55 LFRDDAFPPVSHSLGFEKELG-PNRSKYTGIKMRPELISNPFIYDGAATRDICGALG 113
 QY 72 DCMFLCACAALOKSRHLLDQVTPPGQPSWADDERGSCFCRIQWGFVETDDSLP 131
 DB 114 DCMFLAALAIASLTINEELIARVPPQS--FQENVAGIFHFQWQGEWEVYDDSLP 171
 QY 132 ASRLFCRQREDVWPLPLEKVVAVHGSYTEHMGVADALVDLTGGLAERWNLKGYA 191
 DB 172 DKLIVFVHSAEGSEFMSALLEKAYAKVNSYELALSGCSEAFEDTGGVTEYD----- 227
 QY 192 GSGGQODRPRGHEHRTCRQL--HLKQOCLSCVLSRAGARLGEF-----HAFIVS 243
 DB 228 -----OKAPS-----DLYOIIILKALERSGLACINSIDRLLEITFNLVRGHAYSVT 277
 QY 244 DLRELQAOACQILLRIQNWGRRCQGLMRGEGMSQVDAVASLSQLGGEFVW 303
 DB 278 DAKQVTVYQGVAVNLIRNPMQGVETGKMNDCPSMNTYDPRERDLRVFMEDEGEFWM 336
 QY 304 EEEFLREFDELTVGYVTEAGHLQSLYTERLCHTALPGAWMGQSGAGCRNNSGSPS 363

Db 337 SFDFIREFTKLEICNLPDALKSRLRN-----WNTTYEGTWRRGSTAGGCRN---YPA 389
 QY 364 ----NPKFWLRVSESEV-----YIATLQSRSLHAADWAGRALVGDSTSW 407
 Db 390 TEWNPQKIRLEVDADDYDSRSGCSFLLALAKHR-----RRRRF-----434
 QY 408 SPASIPGKHQVAGLHLKVKERRVNLPRVLSMPVAGTACHAYD-----REVHLRCEL 461
 Db 435 -----GRDMETIGFVAVTQVRELGAQPVHLKRDFFLANASRAQSEHFINLREVSRL 488
 QY 462 SPGYVLAVPSTFLKDPAGEFLLRVPS 487
 Db 489 PGEIIVPSTFEPNKGDFLLRFFS 514

RESULT 6

ID CAN1_HUMAN STANDARD; PRT; 714 AA.
 AC P07384;
 DT 01-APR-1988 (Rel. 07, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Calpain 1, large [catalytic] subunit (EC 3.4.22.17) (Calcium-activated
 DE neutral proteinase) (CAMP) (Mu-type) (muCAMP) (Microfilar-calpain).
 GN CAPN1 OR CAPN1L.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86301172; PubMed=3017764;
 RA Aoki K., Imajoh S., Ohno S., Emori Y., Koike M., Kosaki G., Suzuki K.;
 RT "Complete amino acid sequence of the large subunit of the low-Ca2+-
 RT requiring form of human Ca2+-activated neutral protease (muCAMP)
 RT deduced from its cDNA sequence.";
 RL FEBS Lett. 205:313-317(1986).
 RN [2]

RP SEQUENCE FROM N.A.
 RX MEDLINE=90380278; PubMed=2400579;
 RA Sorimachi H., Ohmi S., Emori Y., Kawasaki H., Saido T.C., Ohno S.,
 RA Minami Y., Suzuki K.;
 RT "A novel member of the calcium-dependent cysteine protease family.";
 RL Biol. Chem. Hoppe-Seyler 371:171-176(1990).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Kidney;
 RL Strausberg R.;
 Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which
 CC catalyze limited proteolysis of substrates involved in
 CC cytoskeletal remodeling and signal transduction.
 CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-I-Xaa, Met-I-Xaa or
 CC Arg-I-Xaa with Leu or Val as the P2 residue.
 CC -!- COFACTOR: Binds 3 calcium ions.
 CC -!- ENZYME REGULATION: Activated by micromolar concentrations of
 CC calcium and inhibited by calpastatin.
 CC -!- SUBUNIT: Heterodimer of a large (catalytic) and a small
 CC (regulatory) subunit.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma
 CC membrane upon Ca++ binding (By similarity).
 CC -!- SIMILARITY: Contains 5 EF-hand calcium-binding domains.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C2; ALSO KNOWN AS THE
 CC CALPAIN FAMILY OF THIOL PROTEASES.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; X04366; CAA27881.1; --
 DR EMBL; BC008751; AA08751.1; --
 DR PIR; A26213; CIRUH.
 DR PIR; S10591; S10591.
 DR HSP; P17655; IDKV.
 DR MEROPS; C02.001; --
 DR MIN; I14220; --
 DR InterPro; IPR002048; EF-hand.
 DR InterPro; IPR001300; Peptidase_C2.
 DR InterPro; IPR000169; Thiolprot_act_site.
 DR Pfam; PF01067; calpain_III; 1.
 DR Pfam; PF00036; ehand; 3.
 DR Pfam; PF00648; Peptidase_C2; 1.
 DR PRINTS; PR00704; CALPAIN.
 DR SMART; SM00230; CysPc; 1.
 DR SMART; SM00034; EFh; 2.
 DR PROSITE; PS00018; EF_HAND; 2.
 DR PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
 DR PROSITE; PS00639; THIOL_PROTEASE_HIS; FALSE NEG.
 DR PROSITE; PS00640; THIOL_PROTEASE_ASN; FALSE NEG.
 KW Hydrolase; Thiol protease; Calcium-binding; Multigene family.
 FT DOMAIN 30 220 THIOL PROTEASE DOMAIN I.
 FT DOMAIN 221 365 THIOL PROTEASE DOMAIN II.
 FT DOMAIN 366 526 DOMAIN III, C2-LIKE DOMAIN.
 FT DOMAIN 527 542 LINKER.
 FT DOMAIN 543 713 DOMAIN IV.
 FT CA_BIND 554 565 EF-HAND 1.
 FT CA_BIND 598 609 EF-HAND 2.
 FT CA_BIND 628 639 EF-HAND 3.
 FT DOMAIN 663 674 ANCESTRAL CALCIUM SITE 4 (POTENTIAL).
 FT DOMAIN 693 704 ANCESTRAL CALCIUM SITE 5 (POTENTIAL).
 FT ACT_SITE 115 115 BY SIMILARITY.
 FT ACT_SITE 272 272 BY SIMILARITY.
 FT ACT_SITE 296 296 BY SIMILARITY.
 FT CONFLICT 548 548 K -> N (IN REF. 3).
 SQ SEQUENCE 714 AA; 81889 MW; 1CB6D7C56D063498 CRC64;

Query Match 16.9%; Score 618.5; DB 1; Length 714;
 Best Local Similarity 32.3%; Pred. No. 1.9e-41;
 Matches 166; Conservative 73; Mismatches 198; Indels 77; Gaps 17;

QY 13 LFRDAAPPAADSSL-FCDLSTPLAOFREDITWRPQECATPRLPDPDPKQVGLG 71
 Db 55 LFRDAEPFPVQSLGKDLG-PNSKTYGIKWRTELLSNPQFIVDGAIRTDICOGALG 113
 QY 72 DCWFLCACAALQKSRHLDDVIPPQPSWADQYRGSTCRWQFGRVVEVTDRLPCL 131
 Db 114 DCWLLAATASTLNDTLHLRVVPHQS--FQNGYAGIFHQLWQFGEWVDVVDLLIK 171
 QY 132 AGRCLFRCQREDVFWLPLEKVKYAKVHSGYEHLMWAGQVADALVDLTGGLAERNLKGVA 191
 Db 172 DGLVVFVHSAEGNEFWSALEKAYAKVNGSYEALSGGSTSEGFEDFTGVTETELRRAP 231
 QY 192 GSGGQDRPGRWEHRTCRQLHLKDDCLISC-----CVLSPPA-CARELGEFHAFIVSDL 245
 Db 232 SLDLYQ-----IILKALRGSLGCSIDISSVLDMEATFKKLVKGHAYSVTGA 279
 QY 246 RELQAGQOCILLRIQNPWGRRCQGLWREGGEGWDAVAASELISQEGEFWVEE 305
 Db 280 KQV-NYRGQVYVLLMRNPGCEVETGAWSDSSSENWNVDPYRQDLRVKMEDEGFWMSP 338
 QY 306 EEFRLREFDELTVGYPTVTEAGHLQSLYTERLLCHTRALPGAWKQSGAGGCRNNSGFFPS-- 363
 Db 339 RDMREERTRLEICNLTPDA--LKSRTIRK--WNTTLYEGTWRRGSTAGGCRN---YPATF 391
 QY 364 --NPKFWLRVSESE-----YIATLQSRSLHAADWAGRALVGDSTSW 409
 Db 392 WNPQKIRLEIDETDDPDYDGRSGCSFVLLALAKHR-----RRRRF-----434
 QY 410 ASIPGKHQVAGLHLKVKERRVNLPRV-LSMPPVAGTACHAYD-----REVHLRCELS 462
 Db 435 ----GRDMETIGFVAVVPELVGQPAVHLKRDFFLANASRAQSEHFINLREVSRL 490

NOTE-Calpain-3 mutations in LGMD2A;
WWW="http://www.dmd.nl/capn3/home.html".

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@sib-sib.ch).

[illegible]

Query Match	16.9%	Score 618;	DB 1;	Length 821;
Best Local Similarity	31.4%;	Pred. No. 2.Se-41;		
Matches 178; Conservative	.79;	Mismatches 194;	Indels 116;	Gaps 19;
QY	13	LFRDAFFPAADSSLEFCDLSPLAQREDITWRBPQETCATPRLFPDDPRSGVKGGLGD	72	
		: :		
Db	74	LYVDPEPPDETSLSFSQRFPI-QF----	VWKRPEICENPRIIDGANTDICQGGELGD	138
		: :		
QY	73	CWFICACAALOKSRHLLDVIPPGQPSWAQDEYRGSTFCRIWGFGRWVEVTDDRFLCLA	132	
		: :		

Db	129	CWFLAAIACITLNOHLLFRVIPHQDS---FIENVAGIFHFQWRKYGEWDVDIDDCLPYTN	180
Qy	133	GRLCFSRCQRDEFWLPLEKVKAYKHGVSYEHLWAGOVADALVDLTGGGLA-----	182
Db	187	NOLVFETKSNNRNEEWSALLEKAYAKLHGYSYALKGCNTTEAMEDFTGGVAEFFEIRDAPS	246
Qy	183	-----ERNLKGA-----SSGGOODRPRWEHTRCQLHLHKDQCCLISCC	223
Db	247	DMYKIMKKAIERGSLMGCSIDDGTNMTYCTSPSGLNMGELIARVMRM-----DNSLLQDS	302
Qy	224	VLSPRAG-----ARELGEFHAFIVSDRLRELGOGAGOCILLIRIONPWG	266
Db	303	DLDRPGSRDERPTTIIPVQYETRMACGLVRGHAYSIGLDEVFFK-GEKVILVRUNPWG	361
Qy	267	RRCHOGLRWREGGWGSQVDAAVASELLSQ--QGEGFVVEEEFLREFDELTYGVYPVTEAG	325
Db	362	QVENWGSWRDKWDKESFDKDEKARLOHQVTEDGEFMWSYEDFIVHFTKLEICNLTAADA-	420
Qy	326	HLOSILYTERLLCHTRAL-PCANWKYGOSAGCCRNSGFP-----SNPKFWLRVSE-----P	374
Db	421	-----LQSDKIQTWTVSNEGWRVGGSGAGCRN---FPDTFTNFQYRLKLEEEDDDPPD	473
Qy	375	SEV-----YIAVLQRSRLHAADMAGRALARLVGDSSHTSWSPASIPGKHYYQAVGLHLMKV---	427
Db	474	SEVICSPVALMKNR-----RKDRKL-----GASLFIGFAIVEVPKE	512
Qy	428	---EKRVNULPRVLSPMPVAGTACHAYDREVHLRCELSPGYLYAVPSTFLKDAPGEFLLR	484
Db	513	MHGKNQHLQKLDFFLYNASKARSATYINMREVSORFRLPPSEYIVPSTVEPHOGEFILR	572
Qy	485	VFSNGR-----VLSLATRAVAKNTT	504
Db	573	VFSEKRNLSEVENTISVDRPVKKKT	599
RESULT 8			
ID	CANL_MOUSE	STANDARD;	PRT; 713 AA.
AC	O35350;		
DT	16-OCT-2001 (Rel. 40, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DE	Calpain I, large [catalytic] subunit (EC 3.4.22.17) (Calcium-activated		
DE	neutrol proteinase) (CAMP) (Mu-type) (muCAMP) (Micromolar-calpain).		
GN	CAPNI OR CAPNI OR CAPAL.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Cf1.		
RA	Poirier C., Poussard S., Faust D.M., Imaizumi-Sherrer T., Weiss M.C.,		
RA	Ducastaing A., Montarras D., Pinset C., Guenet J.-L.;		
RL	Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.		
RL	[2]		
RN	SEQUENCE FROM N.A.		
RC	STRAIN=J29;		
RA	Sahr K.E., Andrabi S., Peters L.L., Chishti A.H.;		
RT	"Cloning and characterization of the cDNA and gene encoding the mouse		
RT	mu-calpain large subunit protein.";		
RL	Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.		
CC	-!- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which		
CC	catalyze limited proteolysis of substrates involved in		
CC	cytoskeletal remodelling and signal transduction.		
CC	-!- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-I-Xaa, Met-I-Xaa or		
CC	Arg-I-Xaa with Leu or Val as the P2 residue.		
CC	-!- COFACTOR: Binds 3 calcium ions (By similarity).		
CC	-!- ENZYME REGULATION: Activated by micromolar concentrations of		
CC	calcium and inhibited by calpastatin.		
CC	-!- SUBUNIT: Heterodimer of a large (catalytic) and a small		
CC	(regulatory) subunit.		
CC	-!- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma		


```

CC      membrane upon Ca++ binding (By similarity).
CC      -1- SIMILARITY: Contains 5 EF-hand calcium-binding domains.
CC      -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C2; ALSO KNOWN AS THE
CC      CALPAIN FAMILY OF THIOL PROTEASES.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; AF021847; AAB72222.1;
DR      EMBL; AF084459; AAC33134.1;
DR      HSSP; P17655; IDKV.
DR      MEROPS; C02.001;
DR      MGD; MGI:88263; Capn1.
DR      InterPro; IPR002048; EF-hand.
DR      InterPro; IPR001300; Peptidase_C2.
DR      Pfam; PF01067; Calpain_III; 1.
DR      Pfam; PF00036; ehfand_3.
DR      Pfam; PF00648; Peptidase_C2; 1.
DR      PRINTS; PR00704; CALPAIN.
DR      SMART; SM00230; Cyspc; 1.
DR      SMART; SM00054; Eph; 2.
DR      PROSITE; PS00018; EF_HAND; 2.
DR      PROSITE; PS00639; THIOL_PROTEASE_CYS; 1.
DR      PROSITE; PS00640; THIOL_PROTEASE_HIS; FALSE NEG.
DR      PROSITE; PS00640; THIOL_PROTEASE_ASN; FALSE NEG.
KW      Hydrolyase; Thiol protease; Calcium-binding; Multigene family.
FT      DOMAIN 30 220
FT      DOMAIN 221 365
FT      DOMAIN 366 525
FT      DOMAIN 526 541
FT      DOMAIN 542 712
FT      CA_BIND 553 564
FT      CA_BIND 597 608
FT      CA_BIND 627 638
FT      DOMAIN 662 673
FT      ACT_SITE 692 703
FT      ACT_SITE 115 115
FT      ACT_SITE 272 272
FT      ACT_SITE 296 296
FT      CONFLICT 488 488
FT      CONFLICT 696 696
SQ      SEQUENCE 713 AA; 82106 MW; 3E1E26C95802B864 CRC64;
Query Match 16.98; Score 615.5; DB 1; Length 713;
Best Local Similarity 32.08; Pred. No. 3.3e-41;
Matches 162; Conservative 80; Mismatches 187; Indels 77; Gaps 16;

```

```

OY      304 EEEFIREDELTYGYPTVEAGHLSLTERLCHTRALPGAWKGSAGCCRNNGSPS 363
DB      337 SFRPIREFTLEICNLTPDALKSRTLN-----NATFEYEGTWKSGTAGCCRN---YPA 389
OY      364 ---NPKWLVSEPSSEV-----YIAVQSRSLAADWAGRALVDSRTSW 407
DB      390 TFWVNPQFKIRLEVDADVDYDNESGCSFLALMOKR-----RRERF----- 434
OY      408 SPASIPKHYQAVGLHMKYKKRVNLPVLSMPVAGTACHAYD-----REYHLRCEL 461
DB      435 -----GRDMEITGFAYQVPRELAGQPVHLKRDFPLANMSRAQSEHINREVSNNRL 488
OY      462 SPGYLAVPSTFLKADAGEFLRYVES 487
DB      489 PGEXIVPSTFEPNKGDFLRFPS 514
RESULT 9
CAN2_RAT
ID      CAN2_RAT STANDARD; PRT; 700 AA.
AC      007009;
DT      01-JUN-1994 (Rel. 29, Created)
DT      01-JUN-1994 (Rel. 29, Last sequence update)
DE      01-MAR-2002 (Rel. 41, Last annotation update)
DE      Calpain 2, large (catalytic) subunit precursor (EC 3.4.22.17)
DE      (Calcium-activated neutral proteinase) (CAMP) (M-type) (M-calpain)
DE      (Millimolar-calpain).
GN      CAPN2.
OS      Rattus norvegicus (Rat).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX      NCBI_TaxID=10116;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE-94032492; PubMed-8218419;
RA      Deluca C.I., Davies P.L., Samis J.A., Elce J.S.;
RT      "Molecular cloning and bacterial expression of cDNA for rat calpain
RT      II 80 kDa subunit.";
RL      Biochim. Biophys. Acta 1216:81-93(1993).
RN      [2]
RP      PARTIAL SEQUENCE.
RX      MEDLINE-21240297; PubMed-11342050;
RA      Moldoveanu T., Hosfield C.M., Jia Z., Elce J.S., Davies P.L.;
RT      "Ca(2+)-induced structural changes in rat m-calpain revealed by
RT      partial proteolysis.";
RL      Biochim. Biophys. Acta 1545:245-254(2001).
RN      [3]
RP      MOTAGENESIS OF LYS-230; LYS-234 AND GLU-504.
RX      MEDLINE-21269273; PubMed-11102442;
RA      Hosfield C.M., Moldoveanu T., Davies P.L., Elce J.S., Jia Z.;
RT      "Calpain mutants with increased Ca2+ sensitivity and implications for
RT      the role of the C(2)-like domain.";
RL      J. Biol. Chem. 276:7404-7407(2001).
RN      [4]
RP      MOTAGENESIS OF CYS-105; HIS-262; ASN-286 AND TRP-288.
RX      MEDLINE-95361909; PubMed-7635186;
RA      Arthur J.S., Gauthier S., Elce J.S.;
RT      "Active site residues in m-calpain: Identification by site-directed
RT      mutagenesis.";
RL      FEBS Lett. 368:397-400(1995).
RN      [5]
RP      X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).
RX      MEDLINE-20069318; PubMed-10601010;
RA      Hosfield C.M., Elce J.S., Davies P.L., Jia Z.;
RT      "Crystal structure of calpain reveals the structural basis for
RT      Ca(2+)-dependent protease activity and a novel mode of enzyme
RT      activation.";
RL      EMBO J. 18:6880-6889(1999).
CC      -1- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which
CC      catalyze limited proteolysis of substrates involved in
CC      cytoskeletal remodelling and signal transduction.
CC      -1- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-I-Xaa, Met-I-Xaa or

```


CC -1- SUBUNIT: Heterodimer of a large (catalytic) and a small (regulatory) subunit.

CC -1- SUBCELLULAR LOCATION: Cytoplasmic.

CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.

CC -1- SIMILARITY: CONTAINS 4 EF-HAND CALCIUM-BINDING DOMAINS.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C2; ALSO KNOWN AS THE CALPAIN FAMILY OF THIOL PROTEASES.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; X92523; CAA63301.1; -

DR EMBL; AF127766; AAD28255.2; -

DR HSSP; P04632; 1DKV.

DR MEROPS; C02.004; -

DR MGD; MGI:107437; Capn3.

DR InterPro: IPR002048; EF-hand.

DR InterPro: IPR001300; Peptidase_C2.

DR InterPro: IPR00169; Thiolprol_act_site.

DR Pfam; PF01067; Calpain_III; 1.

DR Pfam; PF00036; ethand; 3.

DR Pfam; PF00648; Peptidase_C2; 1.

DR PRINTS; PR00704; CALPAIN.

DR SMART; SM00230; Cyspc; 1.

DR SMART; SM00054; Efh; 2.

DR PROSITE; PS00018; EF-HAND; 2.

DR PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.

DR PROSITE; PS00639; THIOL_PROTEASE_HIS; FALSE NEG.

DR PROSITE; PS00640; THIOL_PROTEASE_ASN; FALSE NEG.

DR Hydrolase; Thiol protease; Calcium-binding; multigene family; Repeat;

KM Alternative splicing

FT DOMAIN 49 234

FT DOMAIN 235 428

FT DOMAIN 429 586

FT DOMAIN 587 649

FT DOMAIN 650 820

FT CA_BIND 705 716

FT CA_BIND 735 746

FT DOMAIN 770 781

FT DOMAIN 800 811

FT ACT_SITE 129 129

FT ACT_SITE 334 334

FT ACT_SITE 358 358

FT VARSPLIC 268 315

FT VARSPLIC 595 638

FT VARSPLIC 821 AA; 94213 MW; FF684AE2C4355744 CRC64;

SEQUENCE

Query Match 16.68; Score 605.5; DB 1; Length 821;

Best Local Similarity 29.98; Pred. No. 2.5e-40;

Matches 167; Conservative 83; Mismatches 174; Indels 135; Gaps 18;

QY 13 LFRDAAPPAASSTLFCDSSTPLAOFREDITWRPPEICATRLFPDDREGQVQGLGCD 72

DB 74 LYLDEPEPDTSLFYSGKFPF-QF-----VKRPEIENRPFITIGANRTDTCGGDGD 128

QY 73 CWFICACAAALOKSRHLDDQVTPGQPSWADQYKSGFTCRIMQSGRWVEVTTDRLPCL 132

DB 129 CWFIAAACLTLNRLERLFRVIPHDS--FTENVAGIFHFQWRGRGDWVDVVDCLPYN 186

QY 133 GRLEFSRQGRDVFMLLEKVVAKVGSYHLWAGVADLVLDLGLAERMLK----- 188

DB 187 NQVFTFSNHRNREWSALLERAKVAKLGSGYALKGNTTEMEDFTGGVTEFEIKDAPS 246

QY 189 -----GVAGSG----- 194

DB 247 DMYKMKRAIERGSLMGCSIDDDGNTMNTYGTSPSGLNGLIARVARNNDNSLRLDSIDL 306

QY 195 -GOODRPGWEHRTROLLHKDQCLISCVLSRPAQARELGEPHATVSDRLQAG 253

DB 307 RGSDDR-----RTVPQVETRMAC-----GLVKHAASVGLLEALRK-G 348

QY 254 QCTILLRIQNPWGRNCGWLMREGSGMSQVDAVAASELSQL-QGEFWESEEFREF 312

DB 349 EKVKVLRNPMQGVENSGMSQDKMSFYDDEKARLQHVTEDEFWSTIDFVYHF 408

QY 313 DELTVGPTVPEAGHLSLYERLCHTRAL-PCAMVKGASAGCCRNNGPP---SNPKF 367

DB 409 TKLEICNLTDAD-----LESDKLOTWTVSVNEGFWVGCAGGCRN---PPDFWTPQY 460

QY 368 MLRV-----SEPSEV-----YIAVLRSRRLHADMDAGRAALVGDSTHTSPSPISPKHY 417

DB 461 RLKLEEDDDPEDESEVYICSTFLVALMOKNR-----RKDRKL-----GANL 499

QY 418 QAVGLHLMKY-----EKRRVNLPRVLSMPVAGTACHAVDREYHLCESLPGYTLAVPS 471

DB 500 FTIGFAIYEPKREMGKQKQLOKDFLYNASKARSKTYINRREVSGRFLRPSEYTVPS 559

QY 472 TELKDAPEGFLRFVSTGR 490

DB 560 TYEPHOGEFIRLVFSEKR 578

RESULT 11

CAN2_HUMAN

AC P17655; Q16738;

DT 01-AUG-1990 (Rel. 15, Created)

DT 01-AUG-1990 (Rel. 15, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE Calpain 2, large (catalytic) subunit precursor (EC 3.4.22.17)

DE (Calcium-activated neutral proteinase) (CANP) (M-type) (M-calpain)

DE (Molluscan-calpain).

GN CAPN2 OR CAPNP2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_Taxid=9606;

RN [1]

RP MEDLINE=89166474; PubMed=2852952;

RA Imajoh S., Aoki K., Ohno S., Emori Y., Kawasaki H., Sugihara H., Suzuki K.;

RT "Molecular cloning of the cDNA for the large subunit of the high-Ca2+-requiring form of human Ca2+-activated neutral protease.";

RL Biochemistry 27:8122-8128(1988).

RN [2]

RP SEQUENCE OF 1-79 FROM N.A.

RC TISSUE=Lymph node;

RC MEDLINE=89197947; PubMed=2539381;

RA Hata A., Ohno S., Akita Y., Suzuki K.;

RT "Randomly reiterated negative enhancer-like elements regulate transcription of a human gene for the large subunit of calcium-dependent protease.";

RL J. Biol. Chem. 264:6404-6411(1989).

RN [3]

RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).

RA Strobl S., Fernandez-Catalan C., Braun M., Huber R., Masumoto H., Nakagawa K., Irie A., Sorimachi H., Bouvier G., Bartunik H., Suzuki K., Bode W.;

RT "The crystal structure of calcium-free human m-calpain suggests an electrostatic switch mechanism for activation by calcium.";

RL Proc. Natl. Acad. Sci. U.S.A. 97:588-592(2000).

CC -1- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which catalyze limited proteolysis of substrates involved in cytoskeletal remodeling and signal transduction.

CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-I-Xaa, Met-I-Xaa or Arg-I-Xaa with Leu or Val as the p2 residue.

CC -1- COFACTOR: Binds 3 calcium ions.

DR PROSITE; PS00639; THIOI-PROTEASE_HIS; FALSE_NEG.
DR PROSITE; PS00640; THIOI-PROTEASE_ASN; FALSE_NEG.
KW Hydrolase; Thiol protease; Calcium-binding; Multigene family.
FT CHAIN 1 19 ANCHORS TO THE SMALL SUBUNIT (POTENTIAL).
FT DOMAIN 20 700 CALPAIN 2, LARGE [CATALYTIC] SUBUNIT.
FT DOMAIN 21 210 THIOI-PROTEASE DOMAIN I.
FT DOMAIN 211 355 THIOI-PROTEASE DOMAIN II.
FT DOMAIN 356 514 THIOI-PROTEASE DOMAIN I.
FT DOMAIN 515 529 LINKER.
FT DOMAIN 530 700 DOMAIN IV.
FT CA_BIND 541 552 EF-HAND 1.
FT CA_BIND 585 596 EF-HAND 2.
FT CA_BIND 615 626 EF-HAND 3.
FT DOMAIN 650 661 ANCESTRAL CALCIUM SITE 4 (POTENTIAL).
FT DOMAIN 680 691 ANCESTRAL CALCIUM SITE 5 (POTENTIAL).
FT ACT_SITE 105 105 BY SIMILARITY.
FT ACT_SITE 262 262 BY SIMILARITY.
FT ACT_SITE 286 286 BY SIMILARITY.
SQ SEQUENCE 700 AA; 79228 MW; C3AEDB39CCB56D3B CRC64;

Query Match 16.4%; Score 597.5; DB 1; Length 700;
Best Local Similarity 31.2%; Pred. No. 8.7e-40;
Matches 163; Conservative 81; Mismatches 203; Indels 75; Gaps 17;

QY 1 MRAGRGATPARELFRDAAPDAADSLFCDLSTPLAQFREDITWRPQECATPRLPFPDDP 60
DB 40 LOGGR-----LFHDPSEFAGPAALGYRELGNPKYKGVVWCRTELSCPRFIAGCA 92

QY 61 REGOVKGLGDCWFLCAALOKSHLDVIPPQPSWADOEYKSTCRIMWFGRW 120
DB 93 TRFDICOGALGDCWLLAAASLTLEILARVPRDQ-SFQD-EVAGIFHFQWYGEWV 150

QY 121 EVTTDRPLCLAGRCFSCQREDVFWLPLEKVKVAKVGSVEHLWAGOVADALYDLTGG 180
DB 151 DVVVDRPLTKNGELFLVISAEGSEFWSALKKAYAKLNGSEALSGGTTTGGFEDFTGG 210

QY 181 LAERNWLVAGSGGQDPRGRWHTRCQLLH--LKDQCLISCCVLSPRAG-----AR 232
DB 211 IAEWYEL-----QKAPNLF-----KTIQKALQKSLGCSIDITSAAETAVTSQ 256

QY 233 ELGEFHAFIVSDLRLOGAGQCIILLRIQNPWGRWCQGLRREGGEGSQVDAVASSEL 292
DB 257 KLKVGHAYSVTGAEEVNF--GSIQKILIRNPGEVETGKWNDCNPNWSGVDPEVRRL 315

QY 293 LSQLOGEFWEVEEFREFDELTVG--YPTVEAGHLQSLYTERLLCHTRALPGAWVGQ 350
DB 316 TRRHEDGEFWMFENELRYHSRLCINTLPDPLASDRYKNS--LL-----KLDGNWRGA 369

QY 351 SAGGCRN-NSGSPSPKFWLRYSEPS-----VYIAYLQSLRHLAADWAGRARAL 399
DB 370 TAGGCRNPTFTWNPQYLIKLEEDDEDDPDEGGCTFLIGLIQKHR-----RKQRM 422

QY 400 VGDST-SWSPASIPKHYQAVGLHLWK-----VEKRVNLPVLSMPPVAGTACHAYDRE 454
DB 423 GEDMTIGFIYEVPEFSGQNIHLSKNFNTNAREKSNFTNL-----RE 470

QY 455 VHLRCELSGYLVAPSTLKDAPGEFLRVFSTGRVLSAI 496
DB 471 VLNRRFLPAGEIIVPSFTEPNLNGDFCLRVSEKNANSTVI 512

RESULT 14
CAN3_CHICK
ID CAN3_CHICK STANDARD; PRT; 810 AA.
AC Q92177;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 40, Last annotation update)
DE Calpain 3 large subunit (EC 3.4.22.17) (Calpain L3) (Calpain p94,
DE large [catalytic] subunit) (Calcium-activated neutral proteinase 3)
DE (CANP 3) (Muscle-specific calcium-activated neutral protease 3 large
DE subunit).

GN CAPN3.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=95260862; PubMed=7742367;
RA Sorimachi H., Tsukahara T., Okada-Ban M., Sugita H., Ishiura S.,
RA Suzuki K.;
RT Identification of a third ubiquitous calpain species -- chicken
muscle expresses four distinct calpains.*;
RL Biochim. Biophys. Acta 1261:381-393(1995).
CC -!- FUNCTION: Calcium-regulated non-lysosomal thiol-protease.
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-I-Xaa, Met-I-Xaa or
Arg-I-Xaa with Leu or Val as the P2 residue.
CC -!- ENZYME REGULATION: Activated by micromolar concentrations of
calcium and inhibited by calpastatin.
CC -!- SUBUNIT: Heterodimer of a large (catalytic) and a small
(regulatory) subunit.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic and nuclear (Potential).
CC -!- TISSUE SPECIFICITY: SKELETAL MUSCLE. LOW LEVELS IN SPLEEN,
INTESTINE AND BONE.
CC -!- SIMILARITY: CONTAINS 4 EF-HAND CALCIUM-BINDING DOMAINS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C2; ALSO KNOWN AS THE
CALPAIN FAMILY OF THIOI-PROTEASES.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).

EMBL: D38028; BRA07230.1; --
DR HSSP; P04574; ITALY.
DR MEROPS; C02.004; --
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR001300; Peptidase C2.
DR InterPro; IPR000169; Thiolprot_act_site.
DR Pfam; PF01067; Calpain.III; 1.
DR Pfam; PF00036; ehand; 3; 2.
DR Pfam; PF00648; Peptidase_C2; 1.
DR PRINTS; PR00704; CALPAIN.
DR SMART; SM00230; Cyspc; 1.
DR SMART; SM00054; EFh; 2.
DR PROSITE; PS0018; EF HAND; 2.
DR PROSITE; PS00139; THIOI-PROTEASE_CYS; 1.
DR PROSITE; PS00639; THIOI-PROTEASE_HIS; FALSE_NEG.
DR PROSITE; PS00640; THIOI-PROTEASE_ASN; FALSE_NEG.
KW Hydrolase; Thiol protease; Calcium-binding; Nuclear protein;
KW Multigene family.
FT DOMAIN 43 228 THIOI-PROTEASE DOMAIN I.
FT DOMAIN 229 421 THIOI-PROTEASE DOMAIN II.
FT DOMAIN 422 579 DOMAIN III, C2-LIKE DOMAIN.
FT DOMAIN 580 638 LINKER.
FT DOMAIN 639 809 DOMAIN IV.
FT CA_BIND 694 705 EF-HAND 1 (PROBABLE).
FT CA_BIND 724 735 EF-HAND 2 (PROBABLE).
FT DOMAIN 759 770 ANCESTRAL CALCIUM SITE 3 (POTENTIAL).
FT DOMAIN 789 800 ANCESTRAL CALCIUM SITE 4 (POTENTIAL).
FT ACT_SITE 123 123 BY SIMILARITY.
FT ACT_SITE 327 327 BY SIMILARITY.
FT ACT_SITE 351 351 BY SIMILARITY.
SQ SEQUENCE 810 AA; 93560 MW; E8DE99411C51041E CRC64;

Query Match 16.2%; Score 589.5; DB 1; Length 810;
Best Local Similarity 28.6%; Pred. No. 4.5e-39;
Matches 161; Conservative 87; Mismatches 178; Indels 137; Gaps 19;

RL MOL. Cell. Biol. 15:824-834(1995).
 RN [3]
 RP SEQUENCE FROM N.A. (LONG ISOFORM).
 RC STRAIN-BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Gelniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanalides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Vandeell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers K.-H.C., Blaise R.G., Champagne M., Pfeiffer B.D.,
 RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,
 RA Adair J.F., Abmayyan A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bernier B.P., Bhandari D., Bolshakov S.,
 RA Borovica D., Botchan M.R., Bouck J., Brockstein P., Brotlier P.,
 RA Burris K.C., Busam D.A., Butler H., Cadiot E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doul L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fouts C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwa C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kethum K.A.,
 RA Kimmel B.E., Kodira C.D., Kralt C., Kravitz S., Kulp D., Lai Z.,
 RA Lasro P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy C., Morris J., Moshrefi A.,
 RA Nelson D.R., Nelson K.A., Nixon K., Ruskert D.R., Pached J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Sildenafil K.I., Simpson M., Skipski M.P., Smith T.,
 RA Splet E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svaytshsk R., Tector C., Turner R., Ventler E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of *Drosophila melanogaster*.",
 RL Science 287:2185-2195(2000).
 CC -1- FUNCTION: CALPAINS ARE CALCIUM-ACTIVATED NON-LYSOSOMAL THIO-
 CC PROTEASES. INVOLVED IN THE ORGANIZATION OF THE ACTIN-RELATED
 CC CYTOSKELETON DURING EMBRYOGENESIS.
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-|-Xaa, Met-|-Xaa or
 CC Arg-|-Xaa with Leu or Val as the P2 residue.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS. LONG (SHOWN HERE) AND
 CC SHORT. ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED. HIGHEST IN THE CENTRAL
 CC NERVOUS SYSTEM, BLOOD CELLS AND MIDBUT.
 CC -1- MISCELLANEOUS: THIS PROTEIN BINDS CALCIUM.
 CC -1- SIMILARITY: TO OTHER EF-HAND CALCIUM BINDING PROTEINS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C2, ALSO KNOWN AS THE
 CC CALPAIN FAMILY OF THIOL PROTEASES.
 CC -----
 CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X78555; CAA55298.1; -;
 DR EMBL: X78555; CAA55297.1; ALT_INT.
 DR EMBL: Z46891; CAA86693.1; -;
 DR EMBL: Z46892; CAA86694.1; -;
 DR EMBL: AE003796; AA57563.1; -;
 DR HSSP: P17655; IDKV.
 DR MEROPS: C02.014; -.

DR FlyBase: FBgn0012051; Calpa.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR001300; Peptidase_C2.
DR InterPro: IPR001067; Thiolprot_act_site.
DR Pfam: PF01067; Calpain III; 1.
DR Pfam: PF00036; efhand; 2.
DR Pfam: PF00648; Peptidase_C2; 1.
DR PRINTS: PR00704; CALPAIN.
DR SMART: SM00230; Cyspc; 1.
DR SMART: SM00054; EFh; 2.
DR PROSITE: PS00018; EF_HAND; 1.
DR PROSITE: PS00139; THIOL_PROTEASE_CYS; 1.
DR PROSITE: PS00639; THIOL_PROTEASE_HIS; FALSE_NEG.
DR PROSITE: PS00640; THIOL_PROTEASE_ASN; FALSE_NEG.
KW Hydrolase; Thiol protease; Calcium-binding; Multigene family;
KW Alternative splicing.
FT DOMAIN 68 248 THIOL_PROTEASE DOMAIN I.
FT DOMAIN 249 402 THIOL_PROTEASE DOMAIN II.
FT DOMAIN 403 557 DOMAIN III, C2-LIKE DOMAIN.
FT DOMAIN 558 577 LINKER.
FT DOMAIN 578 827 DOMAIN IV.
FT CA_BIND 712 723 EF-HAND 1 (PROBABLE).
FT CA_BIND 742 753 EF-HAND 2 (PROBABLE).
FT DOMAIN 777 789 ANCESTRAL CALCIUM SITE 3 (POTENTIAL).
FT DOMAIN 807 819 ANCESTRAL CALCIUM SITE 4 (POTENTIAL).
FT ACT_SITE 143 143 BY SIMILARITY.
FT ACT_SITE 299 299 BY SIMILARITY.
FT ACT_SITE 327 327 BY SIMILARITY.
FT VARSPIC 554 558 ENDH -> RTSQ (IN SHORT ISOFORM).
FT VARSPIC 559 828 MISSING (IN SHORT ISOFORM).
FT CONFLICT 35 35 I -> Y (IN REF. 2).
FT CONFLICT 306 306 V -> I (IN REF. 2).
FT CONFLICT 397 397 N -> H (IN REF. 1).
SQ SEQUENCE 828 AA; 93962 MW; 09576D1268BD569C CRC64;

Search completed: September 19, 2002, 14:54:24
Job time: 263 sec

Query Match 16.1%; Score 586; DB 1; Length 828;
Best Local Similarity 30.3%; Pred. No. 8.8e-39;
Matches 167; Conservative 83; Mismatches 185; Indels 116; Gaps 22;
QY 13 LFRDAAPPAADSSIFCDLSTPLAQF--RED--ITWRPQIECATPLRPDDPREGQVKG 68
DB 88 LFEDPLFPASNESL-----QFSRRPDRHIEWLPHIEAENPQFFVEGYSRFDVQ 138
QY 69 LIGDCWFLCACAALQKSRHLLDQVIPPQPSWADQYRGSTCRINWQFGRWVETDDRL 128
DB 139 ELGDCWLLAATANLTQESNLFFRVIPAEQS--FEENYAGIFHFRCYQKWDVVIDRL 196
QY 129 PCLAGRLCFSCQREDVFWLPLEKVTAKYKGSYHLMAGOVADALVLTGGLAERNLK 188
DB 197 PTYNGELMYHSTERNFWSALLEKAYAKLHGSYALKGGSTCEAMEDFTGGVSEWYDLK 256
QY 189 GVAGS-----GGQDRPGRWEHRTCRQLHLKDCGLISCCVLSPRAGAR 232
DB 257 EAPGNLFTILQKAAERNMGCSEPDNPVTEAETPQGLIRHAYSITKVCILD----- 310
QY 233 ELGEFHAFIVSDLELQOGAGQCILLIRTONPWGRRC-WOGLWREGGEGWSQVDAVASE 291
DB 311 -----IVTPNR--QGK---IPMIRMRPNWNEAENWGPWSDDSPWRYIPEQKAE 356
QY 292 LLSOLOGEFWEFEERFERDELTV---GYPTVE---AGHLQ---SLYTERLLCHT 339
DB 357 IGLTFDRDGEFMSQDFLNHFDRVEICNLSPDSLTDQONSGRKRWMSY----- 408
QY 340 RALPGAWKQSGAGGCRN-NSGFPSNPKFWLRVSEPE-----VYIAVLRQSRHAA 390
DB 409 ---EGEWTPGVTAGGCRNFLTDFMHPNYIITLVDPDEDEEGQCIVYVALMOKNR---- 461
QY 391 DWAGRARALVGDHSTWSNPASIPGKHQYQVGLHLMWKEKRRV-NLPRVLSM---PPVAGT 446
DB 462 -----RSKRN-----GMECLTIGFAIYSLNDRLENRPQGLNFRYKSSVGR 504
QY 447 ACHAYD-REVHLRCELSPGYVLAAPPSTFLKDAFGELLRVFSTGRVSL-----AIRAV 499

Db 505 SPHEINTREVCARFKLPKPGHYLIIVPSTFDNPEGEFIRVFSETQNNNEENDHVGYGK 564
QY 500 AKWTPGCAALP 510
Db 565 ADITPGFPPTP 575

Thu Sep 19 15:10:44 2002

us-09-768-877-2.sepl9.rsp

Gencore version 4.5

3010.351 Million cell updates/sec

-
-
-

\$50

ters

• •

500

[illegible]

probable transketolase
probable DNA methylase
brain-specific serine
protein-glutamine
hypothetical protein
connexin 37 - human
protein-glutamine
NADH-ubiquinone oxidoreductase
anti-mullerian hormone
fatty-acid synthase
calpain (EC 3.4.22.25)
hypothetical protein
hypothetical protein
phospholipase A2
anti-mullerian hormone
phospholipase-A(2)

alcium activated neutral
 change 24-Nov-1999
 M.; Suzuki, K.
 se by fusion of genes for
 1; PID:g63333
 ate between those of the
 328/2; 382/1; 438/3; 442
 or arginine residues and
 ology; calpain catalytic d
 otelase; duplication; E
 #status experimental

103;

78; Gaps 16;

REGUNKELLS 71

SECRET

11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 528
 529
 530
 531
 532
 533

[illegible]

Db 165 DGEILFVHSAECTFEWSALLEKAYAKLNGCYESLSGGSTTEGFEFTGGVAEMVDLK --- 221
QY 192 GSGGQODRPGCRWEHRTCRQLLH -----LKQCLISCCVLSPRAGARELGEF-----HAF 240
Db 222 -----RAPRNGHIIIRKALERSLLGCSIDITSADFMEAVTFKKLVKCHAY 267
QY 241 IVSDLREL--QGAGQCILLRTONPWRCWGLNREGGSGVQVDAVASSELLSOLQE 298
Db 268 SVTAFKDVTNGQEQOQ---LIRIRNPGQVEMTGAWSDGSGSEWDNIDPSDREELQLKMD 324
QY 299 GEFWVEEERFLREFDELTVGYPVTEAGHLQSLYTERL-LCHTRALPGAWYKQSGAGGCRN 357
Db 325 GEFWMSFRDPMREFSRLEICNLTPDA-----LTKDELSRWHOTVFEGTWRGSGTAGGCRN 379
QY 358 N-SGFPENPFWRVSEPSSE-----VYIAVLQSRSLHAADWAGRALVGDSTH- 405
Db 380 NPATFWINPQFKILLEDDDDGDEVACSFYVALMQKHR-----RRERRVGGDMHTI 432
QY 406 SWSPASTPGHYQAVGLHLWK-----VEKRVNLPVLSMPVPVAGTACHAYDREVHLRCEL 461
Db 433 GFAYIEVPEEAQSGSNVHLKDKDFLRNQRARSETFINL-----REVSNQIRL 480
QY 462 SPGYLAVPSTELKDPAGEFLLRVES 487
Db 481 PGGEIIVPSTFEPEHKEADFILKVT 506
RESULT 2
CIHUH3
calpain (EC 3.4.22.17) large chain 1 [validated] - human
N:Alternate names: calpain chain L-1; calpain I catalytic chain; low-calcium requiring,
N:Contains: chemotactic factor
C:Species: Homo sapiens (man)
C:Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 08-Dec-2000
C:Accession: A26213; A36740; S10591
R:Aoki, K.; Imaoh, S.; Ohno, S.; Kohke, M.; Kosaki, G.; Suzuki, K.
FEBS Lett. 205, 313-317, 1986
A:Title: Complete amino acid sequence of the large subunit of the low-Ca2+-requiring for
A:Reference number: A26213; MUID:86301172
A:Accession: A26213
A:Molecule type: mRNA
A:Residues: 1-714 <OK>
A:Cross-references: EMBL:X04366; NID:g29663; PIDN:CAA27881.1; PID:g29664
R:Kunimatsu, M.; Higashiyama, S.; Sato, K.; Ohkubo, I.; Sasaki, M.
Biochem. Biophys. Res. Commun. 164, 875-882, 1989
A:Title: Calcium dependent cysteine proteinase is a precursor of a chemotactic factor for
A:Reference number: A36740; MUID:90056492
A:Accession: A36740
A:Molecule type: protein
A:Residues: 2-10 <KUN>
A:Experimental source: erythrocytes
R:Sorimachi, H.; Ohmi, S.; Emori, Y.; Kawasaki, H.; Saido, T.C.; Ohno, S.; Minami, Y.; S
Biol. Chem. Hoppe-Seyler 371(Suppl.), 171-176, 1990
A:Title: A novel member of the calcium-dependent cysteine protease family.
A:Reference number: S10589; MUID:90380278
A:Contents: annotation
A:Note: comparison with other gene products
C:Comment: Calpain I is activated by micromolar concentrations of calcium.
C:Genetics:
A:Gene: GDB:CAPM1; mu-CANP
A:Cross-references: GDB:119749; OMIM:114220
A:Map position: 11pter-11qter
A:Complex: heterodimer of L (large) and S (small) chains
C:Function:
A:Description: catalyzes the hydrolysis of peptides
A:Note: cleaves preferentially after tyrosine, methionine, or arginine residues and before
C:Superfamily: calpain large chain; calmodulin repeat homology; calpain catalytic domain
C:Keywords: acetylated amino end; calcium binding; cysteine proteinase; duplication; EF
F:2-714/Product: calpain large chain 1 #status predicted <NAT>
F:2-10/Product: chemotactic factor #status experimental <CHF>
F:85-337/Domain: calpain catalytic domain homology <CALP>
F:542-573/Domain: calmodulin repeat homology <EF1>
F:585-617/Domain: calmodulin repeat homology <EF2>

F:618-647/Domain: calmodulin repeat homology <EF3>
F:650-682/Domain: calmodulin repeat homology <EF4>
F:683-714/Domain: calmodulin repeat homology <EF5>
F:2/Modified site: acetylated amino end (Ser) (in mature form) #status experimental
F:115,272,296/Active site: Cys, His, Asn #status predicted

Query Match 16.9%; Score 618.5; DB 1; Length 714;
Best Local Similarity 32.3%; Pred. No. 1.5e-40;
Matches 166; Conservative 73; Mismatches 198; Indels 77; Gaps 17;

QY 13 LFRDAAPPAADSSL-FCDLSTPLAQFREDITWRRPQECATPRLFPDPPREGQVKGSLG 71
Db 55 LFRDEAPPPVQPSGLYKDLG-PNSSKTYGIKKRPTTELLSNPQFIVDGAFTDICQCALG 113
QY 72 DCFELCAALAKSRILLDOVIPPQPSWADQYRGSFTCRINQFGRWVZTTDDRLPCL 131
Db 114 DCLLAAIASLTLNDLHLHRVPHGOS--FQNGYAGIFHFQWFGWEDVVVDDLLPIK 171
QY 132 AGRICFSRCOREDFVWLPLEKVKYAKVHGSYEHLMGAVADALVDLTGGLAERWNLKGA 191
Db 172 DGKLVFVHSNEGNEFWSALLEKAYAKVNGSYEALSGSSTSEGFEDTGGVTEYELKAP 231
QY 192 GSGGQODRPGCRWEHRTCRQLLHLDQCLISC-----CVLSPRA-GARELGEFHAIVSDL 245
Db 232 SDLYQ-----IILKALERSLLGCSIDISSVLDMEAITFKLVKCHAYSVTGA 279
QY 246 RELQAGAGCILLRLTONPWRCWGLNREGGSGVQVDAVASSELLSOLQGEFWE 305
Db 280 KQV-NYRGQVSVLIRNPNWGEYETWGAUSSSSNNVDPYERDQLRVKMGDEGFWMSP 338
QY 306 EEFLEFDELTVGYPVTEAGHLQSLYTERLCHTRALPGAWYKQSGAGGCRNNGSPS-- 363
Db 339 RQPRREFTRLEICNLTPDA--LKSRTIRK--WNTTYIETWRRGSTAGGCRN---TPATF 391
QY 364 --NPKFWLRVSEPSSE-----VYIAVLQSRSLHAADWAGRALVGDSTHWSWP 409
Db 392 WYNPQFKIRLDETDDDDDYDRESGCSFVLALMQKHR-----RRRRP----- 434
QY 410 ASIPGKHQYQAVGLHLWKVKRRVNLPRV-LSMPPVAGTACHAYD-----REVHLRCELS 462
Db 435 ---GRDMETIGFVAYTEVPVLPVGPVHLKRDFFFLANASRASEQFINLREYSTRFLP 490
QY 463 PGYILAVPSTFELKDPAGEFLLRVF---STGRVSL 493
Db 491 PGYIVVVPSTFEPEHKEADFILKVT 524

RESULT 3
CIHUH3
calpain (EC 3.4.22.17) large chain 3 - human
N:Alternate names: calpain chain L-3; calpain III catalytic chain; muscle specific,
C:Species: Homo sapiens (man)
C:Date: 28-Apr-1995 #sequence_revision 09-Aug-1997 #text_change 16-Jul-1999
C:Accession: A56218; A34488
R:Richard, I.; Broux, O.; Allamand, V.; Fougereousse, F.; Chiannilkulchai, N.; Bourg,
J.A.; Fardeau, M.; Jackson, C.E.; Cohen, D.; Beckmann, J.S.
Cell 81, 27-40, 1995
A:Title: Mutations in the proteolytic enzyme calpain 3 cause limb-girdle muscular dy
A:Reference number: A56218; MUID:95236448
A:Accession: A56218
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-821 <RIC>
A:Cross-references: GB:X85030; NID:g791039; PIDN:CAA59403.1; PID:g791040
R:Sorimachi, H.; Imaoh-Ohmi, S.; Emori, Y.; Kawasaki, H.; Ohno, S.; Minami, Y.; Suz
J. Biol. Chem. 264, 20106-20111, 1989
A:Title: Molecular cloning of a novel mammalian calcium-dependent protease distinct
A:Reference number: A34688; MUID:90062125
A:Accession: A34488
A:Molecule type: mRNA
A:Residues: 44-445, 'AA', 448-458, 'P', 460-461, 'P', 463-484, 'T', 486-821 <SOR>
C:Genetics:

A:Reference number: A48764; MUID:93374936

A:Accession: A48764

A:Status: preliminary

A:Molecule type: DNA; protein

A:Residues: 1-703 <SOR>

A:Cross-references: GB:DL4479; NID:9441199; PIDN:BAA03370.1; PID:9441200

A:Experimental source: stomach

A>Note: sequence extracted from NCBI backbone (NCBIN:137770, NCBIP:137771)

A:Accession: B48764

A:Status: preliminary

A:Molecule type: DNA; protein

A:Residues: 1-379, 'SS' <SO2>

A:Cross-references: GB:DL4480; NID:9441201; PIDN:BAA03371.1; PID:9495223

A:Experimental source: stomach

A>Note: sequence extracted from NCBI backbone (NCBIN:137773, NCBIP:137775)

C:Superfamily: calpain large chain; calmodulin repeat homology; calpain catalytic domain

C:Keywords: alternative splicing; cysteine proteinase; EF hand; hydrolase

F:75-327/Domain: calpain catalytic domain homology <CALP>

F:532-563/Domain: calmodulin repeat homology <EF1>

F:575-607/Domain: calmodulin repeat homology <EF2>

F:640-672/Domain: calmodulin repeat homology <EF3>

F:105,262,286/Active site: Cys, His, Asn #status predicted

Query Match 16.8%; Score 612; DB 2; Length 703;
Best Local Similarity 31.2%; Pred. No. 4.7e-40;
Matches 155; Conservative 87; Mismatches 197; Indels 58; Gaps 14;

QY 13 LFRDAFPAADSSSL-FCDLSTPLAQFREDITWRPQICATPRLFPDDPREGQVKGGLG 71

DB 45 LFKDEFPACPSALGYKDLG-PGSPDFQIGVWRKPTLCPNPFQVGGATRTDIRQGLG 103

QY 72 DCWFLCACALQKSRHLLDQVIPPQSPWADQVYSGFTCRWQFGRWVETDRLPCL 131

DB 104 DCWLLAAIASLTNEKLLVLPDQSS--FQKDVAGIFHFQWQYGEWVEVDDRLPTK 161

QY 132 AGRLCFSRCQREDVFWLPLEKVKYAKVHGYEHLWAGQVADALVDLTGGLAERNLKV 191

DB 162 NGQLFLHSEEGNEFWLLEKAYAKLVGSGTTEGDFGTFGGISEFYDLK--- 218

QY 192 GSGGQDPRGWEHRTCRQLHLKDKQLISCCV--LSPRAGA-----RELGEFAFVSD 245

DB 219 -----KPPENLYIITQKA--LRKSLGCSIDVSTAAEAATRKLVKGHAYSIVG 269

QY 246 RELQGOAGQCILLRIQNPGRRCWQGLWREGGEGSWQDAVAASELLSQLOEGEFWEE 305

DB 270 EEFNHF-GRPEKLIRLNPNWGEVSWGAWSDNAPWNYIDPRRKEELDKAEDGEFWM 328

QY 306 EEFRLREFDELTVGVPVTEAGHLQSLYTERLLCHTRALPGAWKVGOSAGGCNNSG-FPSN 364

DB 329 SDFLKQYSRLICNLSPDLSSEIHKWNLVFN---GRWTRGAGGCLNPGTYWTN 384

QY 365 PKFWLRSVSEPE-----VYIAVLRSLRHAADWAGRARALVGDSTHS--WSPA 410

DB 385 PQFKIHLDEVEDQEGTSEPCCTVLIGLMOKNR-----RRQKRIQOGMLSIGVAVY 436

QY 411 SIPCKHQVAGLHMKVEKRVNLPVLSMPVAGTACHAYDREHLRCLSPGYILAVP 470

DB 437 QIPRELSHTDAHLGR-----DFFLGRQPSCTCSYTNLREVSRSRLPPQYILVVP 488

QY 471 STFLKDPAGFEFLLRVFS 487

DB 489 STEFFPKDGFCLRVFS 505

RESULT 6
CIHUH2

N:Alternate names: calpain chain 2 - human

C:Species: Homo sapiens (man)

C:Date: 21-Nov-1993 #sequence_revision 09-Aug-1997 #text_change 16-Jul-1999

C:Accession: S10589; A31218; A33529

R:Sorimachi, H.; Ohmi, S.; Emori, Y.; Kawasaki, H.; Saido, T.C.; Ohno, S.; Minami, Y.; S

Biol. Chem. Hoppe-Seyler 371(Suppl.), 171-176, 1990

A:Title: A novel member of the calcium-dependent cysteine protease family.

A:Reference number: S10589; MUID:90380278

A:Accession: S10590

A:Molecule type: mRNA

A:Residues: 1-700 <SOR>

R:Imajoh, S.; Aoki, K.; Ohno, S.; Emori, Y.; Kawasaki, H.; Suzuki, K.

Biochemistry 27, 8122-8128, 1988

A:Title: Molecular cloning of the cDNA for the large subunit of the high-Ca(2+)-requ

A:Reference number: A31218; MUID:89166474

A:Accession: A31218

A:Molecule type: mRNA; protein

A:Residues: 1-210, 'I', 'D', '396-445, 'I', '447-700 <IMA>

A:Cross-references: GB:M23254; NID:9511636; PIDN:AAA35645.1; PID:9511637

A:Note: Parts of this sequence were determined by protein sequencing; the amino end

R:Hata, A.; Ohno, S.; Akita, Y.; Suzuki, K.

J. Biol. Chem. 264, 6404-6411, 1989

A:Title: Tandemly reiterated negative enhancer-like elements regulate transcription

A:Reference number: A33529; MUID:89197947

A:Accession: A33529

A:Molecule type: DNA

A:Residues: 1-67, 'G', '69-72, 'IE', '75-78, 'R' <HAT>

A:Cross-references: DDBJ:J04700; NID:9179910; PIDN:AAA52760.1; PID:9463086

C:Genetics:

A:Gene: GDB:CAPN2; mCAPN; CAPNml

A:Cross-references: GDB:119750; OMIM:114230

A:Map position: lpter-lqter

C:Complex: heterodimer of L (large) and S (small) chains

C:Function:

A:Description: catalyzes the hydrolysis of peptides

A:Note: Cleaves preferentially after tyrosine, methionine, or arginine residues and

C:Superfamily: calpain large chain; calmodulin repeat homology; calpain catalytic do

C:Keywords: acetylated amino end; calcium binding; cysteine proteinase; EF hand; het

F:2-700/Product: calpain large chain 2 #status predicted <MAT>

F:75-327/Domain: calpain catalytic domain homology <CALP>

F:529-560/Domain: calmodulin repeat homology <EF1>

F:572-604/Domain: calmodulin repeat homology <EF2>

F:605-634/Domain: calmodulin repeat homology <EF3>

F:637-669/Domain: calmodulin repeat homology <EF4>

F:2/Modified site: acetylated amino end (Ala) (in mature form) #status predicted

F:105,262,286/Active site: Cys, His, Asn #status predicted

Query Match 16.7%; Score 608; DB 1; Length 700;
Best Local Similarity 30.8%; Pred. No. 9.7e-40;
Matches 159; Conservative 74; Mismatches 209; Indels 74; Gaps 15;

QY 10 ARELFRDAFPAADSSSL-FCDLSTPLAQFREDITWRPQICATPRLFPDDPREGQVKG 68

DB 42 AGTLFODPSFPAIPSAIGFKELG-PYSSKTRGMWRKPTICADPQFIIGGATFTDICQ 100

QY 69 LLGDCWFLCACALQKSRHLLDQVIPPQSPWADQVYSGFTCRWQFGRWVETDRL 128

DB 101 ALGDCWLLAAIASLTNEELIARVPLNQS--FOENYAGIFHFQWQYGEWVEVDDRL 158

QY 129 PCLAGRLCFSRCQREDVFWLPLEKVKYAKVHGYEHLWAGQVADALVDLTGGLAERNL 188

DB 159 PTKDGEILFVHSAEGSEFWLLEKAYAKLVGSGTTEGDFGTFGGIAEWELK 218

QY 189 GYVAGSGQDPRGWEHRTCRQLHLKDKQLISCCVLSPPRAGARELGEF-----HAFIV 242

DB 219 -----KPPNLFKIIQKA--LOKSLGCSIDITSAADSEAITFOKLKVGHAYS 266

QY 243 SDLRELQGOAGQCILLRIQNPGRRCWQGLWREGGEGSWQDAVAASELLSQLOEGEF 302

DB 267 TGAEEVESN-GSLQKILIRINPNWGEVSWGAWSDNAPWNYIDPRRKEELTRRHEDEFW 325

QY 303 VEDEEFLREFDELTVGVPVTEAGHLQSLYTERLLCHT-----RALPGAWKVGOSAGGCR 356

DB 326 MFSDFPLRHYSRL-----EICNLTPDTLTSDTYKKWKLTKMDGNWRRSGTAGGCR 375

QY 357 N-NSGFPSPKFWLRSVSEPE-----VYIAVLRSLRHAADWAGRARALVGDSTH 405

Db 376 NYNTEFWANPOYLKLEEEDEDEDESGCTFLVGLQKR-----RQRKMGEDMT 428
 Oy 406 -SWSPASIRPKHQAVGLHMK-----VEKRRVILPRLSNPVPVAGTACHAVDEVLHRC 460
 Db 429 IGGIIVEVEEELSGQTNLLSKNEFLLNRNRRERSDFEINL-----REVLRK 476
 Oy 461 LSPGYLAVPSTFLKDAPEGFLIRVSTGVSLAI 496
 Db 477 LPEGETIIVPSTFEPKNDGDFCIRFSEKKADYQAV 512

RESULT 7
 B34488
 calpain (EC 3.4.22.17) large chain 3 - rat

N:Alternate names: cysteine proteinase

C:Species: Rattus norvegicus (Norway rat)

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C:Accession: B34488; S10589

R:Sortmach: H.; Imajob-Ohmi, S.; Emori, Y.; Kawasaki, H.; Ohno, S.; Minami, Y.; Suzuki, J. Biol. Chem. 264, 20106-20111, 1989

A:Title: Molecular cloning of a novel mammalian calcium-dependent protease distinct from

A:Reference number: A94688; MUID:90062125

A:Accession: B34488

A:Molecule type: mRNA

A:Residues: 1-821 <SOR>

A:Cross-references: GB:J05121; NID:q205955; PIDN:AAA41790.1; PID:q205956

R:Sortmach: H.; Ohmi, S.; Emori, Y.; Kawasaki, H.; Salido, T.C.; Ohno, S.; Minami, Y.; S

Biol. Chem. Hoppe-Seyler 371(Suppl.), 171-176, 1990

A:Title: A novel member of the calcium-dependent cysteine protease family.

A:Reference number: S10589; MUID:90380278

A:Accession: S10589

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-50, 'E', '52-211, 'V', '213-252, 'K', '254-821 <SO2>

C:Superfamily: calpain large chain; calmodulin repeat homology; calpain catalytic domain

C:Keywords: calcium binding; cysteine proteinase; EF hand; hydrolase

F:649-680/Domain: calpain catalytic domain homology #status atypical <CALP>

F:652-724/Domain: calmodulin repeat homology <EF1>

F:725-754/Domain: calmodulin repeat homology <EF2>

F:757-789/Domain: calmodulin repeat homology <EF3>

F:790-821/Domain: calmodulin repeat homology <EF4>

F:129,334,358/Active site: Cys, His, Asn #status predicted

Query Match 16.4%; Score 599.5; DB 1; Length 821;
 Best Local Similarity 31.0%; Pred. No. 5.6e-39;
 Matches 170; Conservative 78; Mismatches 186; Indels 115; Gaps 18;

Oy 13 LFRDAAPPAADSSLCFCDLSTPLAQRREDITWRPQETICATPRLPPDPREGQVKGGLGD 72
 Db 74 LYPDPEPPDETSLFYEQKFEI-OF-----VWKRPPEICENRFLITIGANKRFDICQGLGD 128
 Oy 73 CWFICACAAALQKSHLLDOYIPGQSPMAOETRGSTCRIOGKRVETVTDRLPCLA 132
 Db 129 CWLLAATACLTLLNERLFRVPHDO--FTENYAGIFHFQFWRGVDVYIDCLPTYN 186
 Oy 133 GRLCFSRCQREDEVFMLPLEKRYAKVGSYEHLMAGQVADALVDLTGLAERMLK----- 188
 Db 187 NQVLFKRSNHRNEMFMSALLKRAYAKLGSTELAKGNTTEMEDTGGVVEFEIKAPPS 246
 Oy 189 -----GVAGSGGQDRPGRWHRTRCROLLHLKDOCLI 220
 Db 247 DMTKIMKALERGLMGCSIDDTNMTYGTSPGIAN--GELLARVRNM-----DNSL 299
 Oy 221 SCCVLSPRAG-----ARLGEHAFIVSDRLQOAGOCILLRIQN 263
 Db 300 RSDLDPRASDDRRPRTIYVQYETRMACGLVGHAVSYGLEALFK-GEKVAVLRN 358
 Oy 264 PMGRRCMOGIMRGSGMSQVDAVAASELISOL-OEGEFVVEEELREDELTVGIPVT 322
 Db 359 PMQOVEMNGSMGDKMDSYVDDEKARLQHOVTEDEGEFMASITDYVTHFTKLEICMTA 418

Oy 323 EAGHLASTYERLLCHTRAL-PCAWYKGSAGCCRRNSGRP-----SNPKFWLAVSE----- 373
 Db 419 DA-----LESCKIQTIVTSVNBGRWAGCSAGCCRN---PPDFEMNPQRYLLLEDDDD 470
 Oy 374 -PSEV-----YIAVLQSRSLAAADWAGARALVGDHSHTSPASIPCKHQAVGLHMKV 427
 Db 471 PDDSEVICSPFLVAMQKNR-----RKDKL-----CANLFTIGFAIYEV 509
 Oy 428 -----EKRRVNLPRVLSNPVAGTACHAVREVLHRCESLPGYLAVPSTFLKDAPEG 481
 Db 510 PKEMGNKHLOKDFFLYNASKARSKTYIMREVSRFLRPEEYIVPSTYEPHOGGEF 569
 Oy 482 LIRVFSTGR 490
 Db 570 ILRVSEKR 578

RESULT 8

S57194
 calpain (EC 3.4.22.17) large chain 2 - chicken

N:Alternate names: m-calpain heavy chain

C:Species: Gallus gallus (chicken)

C:Date: 28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 20-Jun-2000

C:Accession: S57194

R:Sortmach: H.; Tsukahara, T.; Okada-Ban, M.; Sugita, H.; Ishiura, S.; Suzuki, K.

Biochim. Biophys. Acta 1261, 381-393, 1995

A:Title: Identification of a third ubiquitous calpain species - chicken muscle exp

A:Reference number: S57194; MUID:95260862

A:Accession: S57194

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-700 <SOR>

A:Cross-references: EMBL:D38026; NID:q882068; PIDN:BA07228.1; PID:q882069

C:Superfamily: calpain large chain; calmodulin repeat homology; calpain catalytic d

C:Keywords: cysteine proteinase; EF hand; hydrolase

F:75-327/Domain: calpain catalytic domain homology <CALP>

F:529-560/Domain: calmodulin repeat homology <EF1>

F:572-604/Domain: calmodulin repeat homology <EF2>

F:605-634/Domain: calmodulin repeat homology <EF3>

F:637-669/Domain: calmodulin repeat homology <EF4>

F:670-700/Domain: calmodulin repeat homology <EF5>

F:105,262,286/Active site: Cys, His, Asn #status predicted

Query Match 16.4%; Score 597.5; DB 2; Length 700;
 Best Local Similarity 31.2%; Pred. No. 6.5e-39;
 Matches 163; Conservative 81; Mismatches 203; Indels 75; Gaps 17;

Oy 1 MRAGRGATPARELFRDAAPPAADSSLCFCDLSTPLAQRREDITWRPQETICATPRLPPDP 60
 Db 40 LQGR-----LPHDSFSPAGPALGYRELGPMSYKTKGVWCRPTLSCPRINGA 92
 Oy 61 REGVKGGLDCCWFLCACALQKSHLLDOYIPGQSPMAOETRGSTCRIOGKRVGFW 120
 Db 93 TRDTCGALGDDCWLLAALASLTINEELFARVPRD-SPED-ETAGIFHFQMGYEMV 150
 Oy 121 EYTTDRRLPCLAGRLCSRCQREDEVFMLPLEKRYAKVGSYEHLMAGQVADALVDLTGG 180
 Db 151 DVVVYDDRLPTKNGELLFVHSAESESFSALEKRYAKLNGSYELSGTTTEGEDEFDGG 210
 Oy 181 LAERWNLKGVAGSGQODRGRWHRTRCROLLH-LKDOCLISCVLSPRAG-----AR 232
 Db 211 IAEYTEL-----QKAPNLF-----KIQRALQKSLGCSIDTISAETAVTSQ 256
 Oy 233 ELGEHAFIVSDRLQOAGOCILLRIQNPMCRRCMOGIMRGSGMSQVDAVAASEL 292
 Db 257 KLVKGAHYSTGABEVNER-GSIOKLRIRNPGEVEMTKMNDNCNMGVPEVNERL 315
 Oy 293 LSOLQEGEFVVEEELREDELTVG--VPVTEAGHLQSYTERLLCHTRALPGAWYKQ 350
 Db 316 TRRHEGGEFMAFNDFLRHYSRLICNLFPDILASDYKTKMS--LL-----KLGGNMRGA 369
 Oy 351 SAGGCRN-NSGFRSNKRFMLRVSEPS-----YIAVLQSRSLAAADWAGARAL 399

Db 370 TAGGCRNPTNFTWNPOLIKLEEDDDPDDPEGCGTFLGLIQKHR-----RKORKM 422
Qy 400 VGDST-SWSPASIPGKHQYQAVGLRWK-----VEKRVNLPVLSMPVAGTACHAYDRE 454
Db 423 GEDMTIGFAIYVPPFSGGTNIHLKSNFTLNKAREKSNFTNL-----RE 470
Qy 455 VHLRCELSPGYLLAVPTFLKADPGEFLLRVSTGRVLSAI 496
Db 471 VLNRFKLPAAGEYIIVPSTFENLNGDFCLRVSEKKNANSTVI 512

RESULT 9
S57196
calpain (EC 3.4.22.17) large chain 3 - chicken
N:Alternate names: calpain p94 heavy chain; n-calpain-1 heavy chain
C:Species: Gallus gallus (chicken)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
A:Accession: S57196
R:Sorinachi, H.; Tsukahara, T.; Okada-Ban, M.; Sugita, H.; Ishiura, S.; Suzuki, K.
Biochim. Biophys. Acta 1261, 381-393, 1995
A:Title: Identification of a third ubiquitous calpain species - chicken muscle expressed
A:Reference number: S57194; MUID:95260862
A:Accession: S57196
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-810 <SOR>
C:Cross-references: EMBL:D38028; NID:g882072; PIDN:BA07230.1; PID:g1552167
C:Superfamily: calpain large chain; calmodulin repeat homology; calpain catalytic domain
C:Keywords: cysteine proteinase; EF hand; hydrolase
F:93-393/Domain: calpain catalytic domain homology <CALP>
F:746-778/Domain: calmodulin repeat homology <EFH>

Query Match 16.2%; Score 589.5; DB 1; Length 810;
Best Local Similarity 28.6%; Pred. No. 3.3e-38;
Matches 161; Conservative 87; Mismatches 178; Indels 137; Gaps 19;
Qy 13 LFRDAAPADSSLCFSLTFLAQFEDITWRPQETCATPRFPDPREGOVKQGLIGD 72
Db 68 LYEDPDPPTNFTSLFYSQKVP-----KFEWKRPREICENPRTIIGCANRTDICOQELGD 122
Qy 73 CWFCAACALQKSHLLDQVIPPQPSWADQVEYSGFTCRWQFGRVVEVTTDDRLPCLA 132
Db 123 CWFALATLTLNKKLCRVIPHDQS--FTQYAGIFHFQWRYGWDVVDIDCLPTYN 180
Qy 133 GLRCFSRCOREDFWLPLEKVKYAKVHGSYEHLMWAGQVADALVDLTGGLAERNLK 188
Db 181 NLVFTKSSORNEFWALSLEKAKLHGSYALKGGNTTAMEDFTCGVTEFTKIDAPK 240
Qy 189 -----GVAGSG----- 194
Db 241 DIYKIMKHAIARGSLMASSIDDLGPHYGAAPRSDIGELIARMVKLENNAQMTYSTVDYQ 300
Qy 195 GOODRGRWEHRTCOLLHLKQCLISCCVLSPPAGARELGEHAFIVSDRLRQAGQ 254
Db 301 GTERPA-W-----TIMPHQYETRSC-----GLVKHAYSVTAVEETT-YKGE 342
Qy 255 CILLRITQNPWGRRCWGLRREGGWSQVDAVAASLLSOL-QEGEFWVEEEFLREED 313
Db 343 KMLRLVRNPAGQWNGPNSDSSEWNFIDEEEKIRLQHKIAEDGEFVLSLEDFMRHET 402
Qy 314 ELTVGYVPTAGHLQSLYTERLCHTRAL-PGAWVRQSGAGGCRN-NSGFPSPNPKFWLRV 371
Db 403 KLEICNLTPD-----TLEADKLQWTVSVNVEGRVRCSSAGGCRNPDFTWNPYRLKL 457
Qy 372 SEPE-----VYIAVLRSLRHADWAGARALVGDST-SWSPASIP-----GK 415
Db 458 LEEDDDPEDEVICSLFVALMQNR-----RKERKLANLYTIGFAIYVPEPKMHGTY 510
Qy 416 HYQAVGLHLMKVKERR-----VNLPRVLSMPVAGTACHAYDREVHLRCELSPGYLLAVPS 471
Db 511 HHLQKDFELNASKANSKTYINM-----REISERFLRPSEYVILPS 552

Qy 472 TFLKADPGEFLLRVSTGRVLS 494
Db 553 TYEPHQEGEFLLRVSEKR-SLS 574
RESULT 10
A55054
calpain (EC 3.4.22.17) large chain - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 08-Jul-1995 #sequence_revision 03-Aug-1995 #text_change 18-Jun-1999
C:Accession: A55054
R:Emori, Y.; Saigo, K.
J. Biol. Chem. 269, 25137-25142, 1994
A:Title: Calpain localization changes in coordination with actin-related cytoskeleton
A:Reference number: A55054; MUID:95014293
A:Accession: A55054
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-805 <EMO>
A:Cross-references: GB:X78555; NID:g5622387; PIDN:CAA55297.1; PID:g5622288
C:Genetics:
A:Gene: FlyBase:Calpa
A:Cross-references: FlyBase:FBgn0012051
C:Superfamily: calpain large chain; calmodulin repeat homology; calpain catalytic do
C:Keywords: cysteine proteinase; duplication; EF hand; hydrolase
F:50-347/Domain: calpain catalytic domain homology <CALP>
F:576-708/Domain: calmodulin repeat homology <EF1>
F:709-738/Domain: calmodulin repeat homology <EF2>
F:741-773/Domain: calmodulin repeat homology <EF3>
F:774-805/Domain: calmodulin repeat homology <EF4>

Query Match 16.1%; Score 588; DB 1; Length 805;
Best Local Similarity 30.0%; Pred. No. 4.4e-38;
Matches 162; Conservative 84; Mismatches 20; Indels 94; Gaps 18;
Qy 13 LFRDAAPADSSLCFSLTFLAQF--RED--ITWRPQETCATPRFPDPREGOVKQG 68
Db 65 LFEDPLFPASNESL-----QFSRPDRHLEWLPHEIAENPQFVGVSYRFDVQOG 115
Qy 69 LLGDCWFLCACAALQKSHLLDQVIPPQPSWADQVEYSGFTCRWQFGRVVEVTTDDL 128
Db 116 ELGDCWLLAATANLTQESNLFRRVPAEQS--FEENYAGIFHFQWRYGWDVVDIDRL 173
Qy 129 PCLAGLRCSCOREDFWLPLEKVKYAKVHGSYEHLMWAGQVADALVDLTGGLAERNLK 188
Db 174 PTNGELMYHSTKNEFEWSALLEKAYAKLHGSYALKGGSTCEAMEDEFTGGVSEWYDLK 233
Qy 189 GVAGS-----GGQDRPGRWEHRTCOLLHLKQCLISCCVLSPPAGAR 232
Db 234 EAPGNLTILQKAERNMGMCSIEPDPNVTAEPTGLIRGHAYSITKVCILID----- 287
Qy 233 ELGEHAFIVSDRLRQAGQCLILLRIONPWGRR--WGLMRREGGWSQVDAVAASE 291
Db 288 -----IVTPNR--QGG-----IPMIRMPNMGNEAENGPWSDDSPENRYIPEQKAE 333
Qy 292 L-LSOIQEGEFWVEEEFLREDFELTVGYVPTAGHLQSLYTERLCHTRALPGAWVKQ 350
Db 334 IGLTDRDGEFWMSFQDLNHFDRVEICNLSPDSLTDQHQHSGKRWKWSMYEGEWTTPGV 393
Qy 351 SAGGCRN-NSGFPSPNPKFWLRVSEPE-----VYIAVLRSLRHADWAGARALVG 401
Db 394 TAGGCRNFLTWFHNPQYIITVDPDDEEGCTVIVALMQNR-----RSKRNM- 444
Qy 402 DSHTSWSPASIPGKHQYQAVGLHLMKVKERR--NLPRVLSM---PPVAGTACHAYD-REVEH 456
Db 445 -----GMECLTIGFAIYLSNDRLENRPGQLNFRYKSSVGRSPHINTREVC 492
Qy 457 LRCELSPGYLLAVPTFLKADPGEFLLRVSTGRVLS-----AIRAVAKNTTTPGAALP 510
Db 493 ARFKLPFGHYLIIVPSTFDPNEEGEFTIRVSETQNNEENDHVGYGKADTTTPGPTP 552

RESULT 11

S44749
 C06G4.2 protein - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C:Accession: S44749
 R:Waterston, R.
 submitted to the EMBL Data Library, November 1993
 A:Description: Sequence of the C. elegans cosmid C06G4.
 A:Reference number: S44747
 A:Accession: S44749
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-653 <WAT>
 A:Cross-references: EMBL:L25598; NID:g409290; PIDN:AAA27940.1; PID:g409293
 C:Genetics:
 A:Introns: 26/1; 216/3; 269/3; 357/2; 401/2; 558/1; 583/3; 613/3
 C:Superfamily: C06G4.2 protein; calpain catalytic domain homology
 F:245-498/Domain: calpain catalytic domain homology <CALP>

Query Match 15.9%; Score 579; DB 1; Length 653;
 Best Local Similarity 30.8%; Pred. No. 1.7e-37;

Matches 157; Conservative 69; Mismatches 176; Indels 108; Gaps 16;

```

OY 10 ARELFRDAFPAADSSLRCDLSTPLAQRFDITWRKRPDITCAPPLFPDDPREGOVKGL 69
DB 217 SRLLEDPOFLANDSSLEFSKRP-----KREVMLEPGITREPOLITEGHSRFYIOGE 271
OY 70 LGDCWFLCACAALQKSRHLIDVIPPQGSFADQERSGFQRIQGFQWVEYTTDDRLP 129
DB 272 LGDCWFLCACAALQKSRHLIDVIPPQGSFADQERSGFQRIQGFQWVEYTTDDRLP 129
OY 130 CLAGRLCFSRCQREDVFWLPLEKRYAVKVSYEHLMAQOVADALVDLTGLAERWNLK- 188
DB 330 TSNGLTLYMHSANSEFALLEKAYAKLFGSYEALKGSTSEALQEDMGTGLTFLDKN 389
OY 189 -----GVAAGSGGODRGRWRHRTCRQLHLKQCLISCCVLSPPAGAR 232
DB 390 PPRNIMOMMNGEFGSIFGCSIEAD-PNVMWAKMSNGLY----- 428
OY 233 ELGEHFIVSDRLAQAGQ-CILLRIONPNC-RRCOMGLMREGGSGQVDAVAS 290
DB 429 ---KGHAISTIGCRIVDPNGOTCI-LRINPMGNEBOMGPNWSDNSREMSVSDSVKQ 483
OY 291 EL-LSQLOGEFWEVEEFLREDELTV--GYPV-TEAGHLQSLYTERLCHTRALPGA 345
DB 484 DMGLKTFHDGGEFMSFDFPMNFEMECINLGPVADQVYQMTGVKAAGMVAANTHDGA 543
OY 346 WVKQSGAGGCGNN--SGEPPNPKFWLVSF--SEYIYAVLOSRLHADMAGRARAL 399
DB 544 WVKQSGAGGCGNNYITFANNPQFRVQLTSDPDDELCTAGNNNGRLSKOFFFAANKSAM 603
OY 400 VGDSTHSWSPASIPGKHQYAVGLHLMKYEKRRVNLPRVLSMPVAGTACHADRVHLRC 459
DB 604 RSAAF-----INL-----RMKTRGF 618
OY 460 ELSPGYIYAVSTELKADAGEFLLRVESTG 489
DB 619 RVPGNVYVVPSTEEPNDEAEFMLRVYTNQ 648

```

RESULT 12

JC5772
 tissue-specific calpain htra-3 - human
 C:Species: Homo sapiens (man)
 C:Date: 24-Jan-1998 #sequence_revision 13-Mar-1998 #text_change 18-Jun-1999
 C:Accession: JC5772
 R:Mugita, N.; Kimura, Y.; Ogawa, M.; Sawa, H.; Nakao, M.
 Biochem. Biophys. Res. Commun. 239, 845-850, 1997
 A:Title: Identification of a novel, tissue-specific calpain htra-3; a human homologue of
 A:Reference number: JC5772; MUID:98042481

A:Accession: JC5772
 A:Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-647 <MUG>

A:Cross-references: GB:U94346
 C:Comment: This protein acts as a cellular receptor of calcium ions in physiologic
 C:Genetics:

A:Gene: htra-3
 A:Map Position: 11q14

C:Superfamily: human tissue-specific calpain htra-3; calpain catalytic domain homo
 F:1-48/Domain: I #status predicted <IND>

F:49-309/Domain: cysteine protease #status predicted <CPT>
 F:51-329/Domain: calpain catalytic domain homology <CALP>

F:310-505/Domain: III #status predicted <III>
 F:317-334/Domain: calcium binding #status predicted <CAL>

F:506-639/Domain: T #status predicted <TDD>

Query Match 15.6%; Score 569.5; DB 2; Length 647;
 Best Local Similarity 30.9%; Pred. No. 9.4e-37;

Matches 159; Conservative 76; Mismatches 196; Indels 83; Gaps 18;

```

OY 13 LFRDAFPAADSSLRCDLSTPLAQRFDITWRKRPDITCAPPLFPDDPREGOVKGLD 72
DB 26 LREDPLEPATDLSLYK-GTPGPAYR---WKRPKICDPRLFYDGISSHDLQGVGN 80
OY 73 CWFICACAALQKSRHLIDVIPPQGSFADQERSGFQRIQGFQWVEYTT 124
DB 81 CWFVACSSLASRESLMQVY----PDWKEQEDPRKAQAYAGIRHFHWRIG-MVDVYI 135
OY 125 DRLPLCLAGRLCFSRCQREDVFWLPLEKRYAVKVSYEHLMAQOVADALVDLTGLAER 184
DB 136 DERLPLTVNQMLYCHSNSNEFEFCALYERAKYAKIAGCYQALDGMVADALVDFGVSER 195
OY 185 WNLKGVAGSGGODRGRWRHRTCRQLHLKQCLISCCVLSPPAGARL-----G 235
DB 196 IDLREGDFANDTCKRQLEF---RLKAVHSRGGLISASIKAVTADMARLACGLYK 251
OY 236 EHFHFIVSDRLAQAGQ-----GCILLRIONPNC-RRCOMGLMREGGSGQVDA 287
DB 252 THAVAVTDVRYKRVLLTHLTLAFKSEKIDMLRLNPMGRMNGPMSPTSEKQVSKS 311
OY 288 VASELISQLO-EGEWEVEEFLREDELTVGYVTEAGHLQSLYTERLCH--TRAL 342
DB 312 EREKGVTVQDDGEFWMGECVCR-----YFTDILIKRVINTSHLSIKTWPEARL 363
OY 343 PGAVY-----KGQSGGCGNN--NSGPPNPKFWLVSF--SEYIYAVLOSRLHAD- 393
DB 364 HGAVTLHEDPRKNGGCGCINHKOTGF-QNPQYIFEVKRPEDVLLCICDQRPSTRERE 422
OY 394 GRARLVDSHTSWSPASIPGKHQYAVGLHLMKYEKRRVNLPRVLSMPVAGTACHAD 453
DB 423 GGENYL-----AIGPDIYKVENNOY--RMSLQKKAASSIYINSR 461
OY 454 EYHLRCELSPGYIYAVSTELKADAGEFLLRVESTG 489
DB 462 SVFLRTDQEGRYVITPTFEPGHGFEFLRVFT 495

```

RESULT 13

S71885
 sex-determining protein transformer-3 - *Caenorhabditis elegans*
 N:Alternate names: tra-3 protein
 C:Species: *Caenorhabditis elegans*
 C:Date: 29-Jan-1998 #sequence_revision 06-Feb-1998 #text_change 29-Oct-1999
 C:Accession: S71885; T23631
 R:Barnes, T.M.; Hodgkin, J.
 EMBO J. 15, 4477-4484, 1996
 A:Title: The tra-3 sex determination gene of *Caenorhabditis elegans* encodes a membe
 A:Reference number: S71885; MUID:97042339
 A:Accession: S71885
 A:Status: nucleic acid sequence not shown
 A:Molecule type: DNA

[illegible]

RESULT 14
A39343
calpain (EC 3.4.22.17) large chain - fluke (Schistosoma mansoni)
C:Species: Schistosoma mansoni
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C:Accession: A39343; A45642
 R:Andresen, K.; Tom, T.D.; Strand, M.
 J. Biol. Chem. 266, 15085-15090, 1991
 A:Title: Characterization of cDNA clones encoding a novel calcium-activated neutral
 A:Reference number: A39343; MUID:91332027
 A:Accession: A39343
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-758 <AND>
 A:Cross-references: GB:M67499; NID:gl60936; PIDN:AAA29858.1; PID:gl60937
 R:Karcz, S.R.; Podesta, R.B.; Siddiqui, A.A.; Dekaban, G.A.; Strejan, G.H.; Clarke,
 Mol. Biochem. Parasitol. 49, 333-336, 1991
 A:Title: Molecular cloning and sequence analysis of a calcium-activated neutral prot
 A:Reference number: A45642; MUID:92131071
 A:Accession: A45642
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-14, 'A', 16-120, 'A', 121-127, 'H', 129-213, 'R', 215-384, 'VTC', 388-440, 'S', 44
 A:Cross-references: GB:M74233; NID:gl60934
 A:Note: sequence extracted from NCBI backbone (NCBTRP:79194)
 C:Superfamily: calpain large chain; calmodulin repeat homology; calpain catalytic dc
 C:Keywords: calcium binding; cysteine proteinase; duplication; EF hand; hydrolase
 F:124-380/Domain: calpain catalytic domain homology <CALP>
 F:628-660/Domain: calpain repeat homology <EF1>
 F:661-690/Domain: calmodulin repeat homology <EF2>
 F:693-726/Domain: calmodulin repeat homology <EF3>
 F:727-758/Domain: calmodulin repeat homology <EF4>
 F:154,313,337/Active site: Cys, His, Asn #status predicted

Query Match 14.8% Score 540; DB 1; Length 758;
 Best Local Similarity 30.0% Pred. No. 2.4e-34;
 Matches 155; Conservative 77; Mismatches 219; Indels 66; Gaps 19;

Query Match 14.8%; Score 540; DB 1; Length 758;
Best Local Similarity 30.0%; Pred. No. 2.4e-34;
Matches 155; Conservative 77; Mismatches 219; Indels 66; Gaps

QY	11	RELFRDAAFPAAOSSULFCDLSTPLAQFREDITWRRPOEICATRLRFPDDPREGQVKQLL	70
Db	97	RTLWEDPDPFANDKAI-----GNLPDRFERIEWKRLNPNKAFPAGASRFDEQAGL	151
QY	71	GDCWFLCACAAALKSRLHLDVIPPQGPWSAQDEYRGSFTCRWQGRWVEVTTDRLPC	130
Db	152	GDCWLLAVVASISGYQLFQDVVPKDO-ELKGPYYGVVFRFWRGHWVEVLIDRLPV	210
QY	131	LAG--RLCFSRCQREDVFWLPLEKYYKVHGSYEHLMAGQAVDALVDVTGLGAERNLK	188
Db	211	ROGLNTLVFWHMSNDPTFEWSALLEKAYAKLNGCYAHLSGSGSSEAMEDLTGGTCLSEL-	269
QY	189	GVAGSGGQODRPRGWEHRTCROLLHLKDO-----CLISC---CVLSPRAGARELGE	236
Db	270	-----NOREPS-----DLIDQLKIYAQRCLMGCSIDSSWEQXMDNGLGS	312
QY	237	FHAFVSDURELQGOAGQCILLRIQNPWG--RRCWGLWREGGEQWSQYDAVAASEL-LS	294
Db	313	-HAYSTGVYVP--NYRGOTQWLRLNPNMGDSHEWKGACDGPQWREISEQEKKNILNS	370
QY	295	QLOGEFWEHEBEFLREFEDELTVGPVTEA-GHLOSXYTERLLCHTRALPGWAVKQSGAS	353
Db	371	FTADGEFWSYEDFCYLFSEVVECHLGLESLEYNQNFHKRRL-DEATFSGQMQRNVAG	429
QY	354	CGRRN--SGPPSNPKFWLRSEFSEVYIAVLQRSRLHAADWAGRARALVGDSTSWPASI	412
Db	430	GCINNRTTYNTWQFRITVEDPD-----PDDDNKC SVLIGLMQTDIRKK--	474
QY	413	PKRHYQAVGLHLHKVE---KRRVNLPRVLSMPVPVAGTACHAYDREVHLRCELSPGYLAY	469
Db	475	VGADFPQIGFWYNADDDLNTLLSRQALLTRSPIAKSQ-FINTREYTAQFRVPPGYSVVI	533
QY	470	PSFLDKAFGEFLLRVFSRGVSLSAIRAVAKNTTPG	506
Db	534	PSFEDNIEVNFLLRVFS--OTSITEDELDENTNOG	568

RESULT 15
S57195

us-09-768-877-2.sepl9.raii

GenCore version 4.5
(c) 1993 - 2000 Com

```
; Search time 14.68 Seconds
(without alignments)
1118.120 Million cell updates/sec
```

...SQEMLGOLFQEVSMVMKT 672

5

ers: 231628

```
OMB.pep:*  
OMB.pep:*  
OMB.pep:*  
OMB.pep:*  
S_COMB.pep:*  
files1.pep:*
```

ected by chance to have a
of the result being printed,
score distribution.

	Description
-2	Sequence 2, Appl
-18	Sequence 18, Appl
-6	Sequence 6, Appl
-4	Sequence 4, Appl
8	Sequence 8, Appl
-10	Sequence 10, Appl
-12	Sequence 12, Appl
A-1	Sequence 1, Appl
-1	Sequence 1, Appl
A-2	Sequence 2, Appl
-2	Sequence 2, Appl
-22	Sequence 22, Appl
-24	Sequence 24, Appl
-27	Sequence 27, Appl
-7	Sequence 7, Appl
A-7	Sequence 7, Appl
-23	Sequence 23, Appl
-7	Sequence 7, Appl
-25	Sequence 25, Appl
-28	Sequence 28, Appl
-14	Sequence 14, Appl
-26	Sequence 26, Appl
-16	Sequence 16, Appl
-17	Sequence 17, Appl
-18	Sequence 18, Appl
-97	Sequence 97, Appl

Sequence 97	Appb
Sequence 95	Appb
Sequence 95	Appb
Sequence 98	Appb
Sequence 98	Appb
Sequence 99	Appb
Sequence 99	Appb
Sequence 96	Appb
Sequence 96	Appb
Sequence 94	Appb
Sequence 94	Appb
Sequence 4	Appb
Sequence 9	Appb
Sequence 3	Appb
Sequence 2	Appb
Sequence 2	Appb
Sequence 10	Appb

h 672;

ATPPLFPDDP	60
ATPRLFPDDP	60
CRIMQGRMV	120
CRIMQGRMV	120
DALVDLTGG	180
DALVDLTGG	180
RELGEFHAF	240
RELGEFHAF	240

Db 241 IVSDRLRQAGOCILLRLRQNPWRCWGLWREGGEGSWQVDAVASSELLSOLQGE 300
Qy 301 FWVEEEFLREFDELTVGYPVTEAGHLQSLYTERLLCHTRALPGAWKVGQSGAGGCRNNSG 360
Db 301 FWVEEEFLREFDELTVGYPVTEAGHLQSLYTERLLCHTRALPGAWKVGQSGAGGCRNNSG 360
Qy 361 FPSNPKFWLRVSESEVYIAVQSRRLHAADWAGRARALVGDSTHSWSPASIPGKHQYQAV 420
Db 361 FPSNPKFWLRVSESEVYIAVQSRRLHAADWAGRARALVGDSTHSWSPASIPGKHQYQAV 420
Qy 421 GLHLWKVKRRVNLPRVLSMPVAGTACHAYDREHLCESLSPGYLLAVPSTFLKADPGE 480
Db 421 GLHLWKVKRRVNLPRVLSMPVAGTACHAYDREHLCESLSPGYLLAVPSTFLKADPGE 480
Qy 481 FLLRVFSTGRVSLSAIRAVAKNTTPGAALPAGEMGTVOLGRSWRVGOTAGGSRNPFASPT 540
Db 481 FLLRVFSTGRVSLSAIRAVAKNTTPGAALPAGEMGTVOLGRSWRVGOTAGGSRNPFASPT 540
Qy 541 NPCPFSVPEGPGRCVRITLHQCRLSDTEFHPHIGFHFQVPEGGRSQDAPPLLLQBP 600
Db 541 NPCPFSVPEGPGRCVRITLHQCRLSDTEFHPHIGFHFQVPEGGRSQDAPPLLLQBP 600
Qy 601 LSCVPHYAQVSRCLLPAGTYKVPSTYLPDTEGAFVTIATRIDRPSIHSQEMLGQF 660
Db 601 LSCVPHYAQVSRCLLPAGTYKVPSTYLPDTEGAFVTIATRIDRPSIHSQEMLGQF 660
Qy 661 LQEVSMVMKMT 672
Db 661 LQEVSMVMKMT 672

RESULT 2
US-09-422-869-18
; Sequence 18, Application US/09422869
; Patent No. 6235481
; GENERAL INFORMATION:
; APPLICANT: POLONSKY, KENNETH S.
; APPLICANT: HORIKAWA, YUKIO
; APPLICANT: ODA, NAOHISA
; APPLICANT: COX, NANCY J.
; APPLICANT: SREENAN, SEAMUS
; APPLICANT: ZHOU, YUN-PING
; APPLICANT: OTANI, KENICHI
; APPLICANT: HANIS, CRAIG L.
; APPLICANT: BELL, GRAEME I.
; TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
; FILE REFERENCE: ARCD:307
; CURRENT APPLICATION NUMBER: US/09/422,869
; CURRENT FILING DATE: 1999-10-21
; EARLIER APPLICATION NUMBER: 60/134,175
; EARLIER FILING DATE: 1999-05-13
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 666
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-422-869-18

Query Match 81.2%; Score 2963; DB 4; Length 666;
Best Local Similarity 81.1%; Pred. No. 3.7e-289;
Matches 544; Conservative 45; Mismatches 76; Indels 6; Gaps 2;
Qy 1 MRAGRGATPAELFRODAAPFADSSLLFCLSTPLAOPREDITWRRPQECATPRLPDDP 60
Db 1 MRVRAETPAELFRODAAPFADSSLLFCLSTPLAOPREDITWRRPQECATPQLPDPNP 60
Qy 61 REGOVKGLGDCWFLCACAAKQSRHLLDQVIPPQGPSWADQYGRSFTCRWQFGRW 120
Db 61 WEGOVKGLGDCWFLCACAAKQSRHLLDQVIPPQGPSWADQYGRSFTCRWQFGRW 120

Qy 121 EVTTDRPLCLAGRLCFSRCQREDVFWLPLEKVTAKVHGSYHLWAGOVADALYDLTGG 180
Db 121 EVTTDRPLCLAGRLCFSRCQREDVFWLPLEKVTAKVHGSYHLWAGOVADALYDLTGG 180
Qy 181 LAERNLKGAGSGGQDRPGWEHRTCRQLHLKDOCLISCCVLSPRAGARELGEFRAF 240
Db 181 LAERNLKGAGSGGQDRPGWEHRTCRQLHLKDOCLISCCVLSPRAGARELGEFRAF 240
Qy 241 IVSDRLRQAGOCILLRLRQNPWRCWGLWREGGEGSWQVDAVASSELLSOLQGE 300
Db 241 IVSDRLRQAGOCILLRLRQNPWRCWGLWREGGEGSWQVDAVASSELLSOLQGE 300
Qy 301 FWVEEEFLREFDELTVGYPVTEAGHLQSLYTERLLCHTRALPGAWKVGQSGAGGCRNNSG 360
Db 301 FWVEEEFLREFDELTVGYPVTEAGHLQSLYTERLLCHTRALPGAWKVGQSGAGGCRNNSG 360
Qy 361 FPSNPKFWLRVSESEVYIAVQSRRLHAADWAGRARALVGDSTHSWSPASIPGKHQYQAV 420
Db 361 FPSNPKFWLRVSESEVYIAVQSRRLHAADWAGRARALVGDSTHSWSPASIPGKHQYQAV 420
Qy 421 GLHLWKVKRRVNLPRVLSMPVAGTACHAYDREHLCESLSPGYLLAVPSTFLKADPGE 480
Db 421 GLHLWKVKRRVNLPRVLSMPVAGTACHAYDREHLCESLSPGYLLAVPSTFLKADPGE 480
Qy 481 FLLRVFSTGRVSLSAIRAVAKNTTPGAALPAGEMGTVOLGRSWRVGOTAGGSRNPFASPT 540
Db 481 FLLRVFSTGRVSLSAIRAVAKNTTPGAALPAGEMGTVOLGRSWRVGOTAGGSRNPFASPT 540
Qy 541 NPCPFSVPEGPGRCVRITLHQCRLSDTEFHPHIGFHFQVPEGGRSQDAPPLLLQBP 600
Db 541 NPCPFSVPEGPGRCVRITLHQCRLSDTEFHPHIGFHFQVPEGGRSQDAPPLLLQBP 600
Qy 601 LSCVPHYAQVSRCLLPAGTYKVPSTYLPDTEGAFVTIATRIDRPSIHSQEMLGQF 660
Db 601 LSCVPHYAQVSRCLLPAGTYKVPSTYLPDTEGAFVTIATRIDRPSIHSQEMLGQF 660
Qy 661 LQEVSMVMKMT 671
Db 661 LQEVSMVMKMT 665

RESULT 3
US-09-422-869-6
; Sequence 6, Application US/09422869
; Patent No. 6235481
; GENERAL INFORMATION:
; APPLICANT: POLONSKY, KENNETH S.
; APPLICANT: HORIKAWA, YUKIO
; APPLICANT: ODA, NAOHISA
; APPLICANT: COX, NANCY J.
; APPLICANT: SREENAN, SEAMUS
; APPLICANT: ZHOU, YUN-PING
; APPLICANT: OTANI, KENICHI
; APPLICANT: HANIS, CRAIG L.
; APPLICANT: BELL, GRAEME I.
; TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
; FILE REFERENCE: ARCD:307
; CURRENT APPLICATION NUMBER: US/09/422,869
; CURRENT FILING DATE: 1999-10-21
; EARLIER APPLICATION NUMBER: 60/134,175
; EARLIER FILING DATE: 1999-05-13
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 517
; TYPE: PRT
; ORGANISM: Human
US-09-422-869-6

Query Match 74.4%; Score 2715.5; DB 4; Length 517;
Best Local Similarity 76.8%; Pred. No. 2e-264;
Matches 516; Conservative 1; Mismatches 0; Indels 155; Gaps 1;

```

QY 1 MRAGGATPARELFRDAAFPAAADSLFCDSLTPLAOFREDITWRRPOICATPRLFPDDP 60
Db 1 MRAGGATPARELFRDAAFPAAADSLFCDSLTPLAOFREDITWRRPOICATPRLFPDDP 60
QY 61 REGOVKOGLLGDCWFLCACAALQKSRHLDDQVIPPQPSMADEYRGSFTCRIMQFGRWV 120
Db 61 REGOVKOGLLGDCWFLCACAALQKSRHLDDQVIPPQPSMADEYRGSFTCRIMQFGRWV 120
QY 121 EYTTDDRLPCLAGRLCFSCQREDVEFWLPLEKYAFAVHGSYEHLMAGVADALVDTLTCG 180
Db 121 EYTTDDRLPCLAGRLCFSCQREDVEFWLPLEKYAFAVHGSYEHLMAGVADALVDTLTCG 180
QY 181 LAERMLKGVASGGGQODRGREHRTCRQLHLKQCLISCCVLSFRAGARELGEFHAF 240
Db 181 LAERMLKGVASGGGQODRGREHRTCRQLHLKQCLISCCVLSFRAGARELGEFHAF 240
QY 241 IYSDLELQAGOCILLRIQNPWGRRCMOGLMREGEGMSQVDAVAASELSQLQEGE 300
Db 241 IYSDLELQAGOCILLRIQNPWGRRCMOGLMREGEGMSQVDAVAASELSQLQEGE 300
QY 301 FWEEBEFLREFDELTVGYPVTEAGHLSYTERLLCHTRALPGAWKQGSAGCRRNSG 360
Db 301 FWEEBEFLREFDELTVGYPVTEAGHLSYTERLLCHTRALPGAWKQGSAGCRRNSG 360
QY 361 FPSNPKFMLRVSPSEYITAVLQSRHLHADNAGARALVGDSTMSPASITGKHYQAV 420
Db 361 FPSNPKFMLRVSPSEYITAVLQSRHLHADNAGARALVGDSTMSPASITGKHYQAV 420
QY 421 GLHLMVEKRRVNLPRVLSMPVAGTACHAYDREVHLRCELSPGYTLAVPSTFLKDAPE 480
Db 421 GLHLMVEKRRVNLPRVLSMPVAGTACHAYDREVHLRCELSPGYTLAVPSTFLKDAPE 480
QY 481 FLIRVSTGRVSLRAIRAVAKTTPGALPAGEMGTVOLGSMRWGOTAGSRNFPASIPT 540
Db 481 FLIRVSTGRVSLRAIRAVAKTTPGALPAGEMGTVOLGSMRWGOTAGSRNFPASIPT 540
QY 541 NRCFPEVSGPGRVCVRITLHQHCRSPDTEPHIGFHLFOVPEGRSODAPPLLOEPL 600
Db 541 NRCFPEVSGPGRVCVRITLHQHCRSPDTEPHIGFHLFOVPEGRSODAPPLLOEPL 600
QY 601 LSCVPHRYAOEVSRLCLTPAGTYKVPSTYLPDTGCAFTVTATRDPSIHISOEMLQOF 660
Db 601 LSCVPHRYAOEVSRLCLTPAGTYKVPSTYLPDTGCAFTVTATRDPSIHISOEMLQOF 660
QY 661 LOEVSVMAMVKT 672
Db 661 LOEVSVMAMVKT 672
QY 506 LOEVSVMAMVKT 517
Db 506 LOEVSVMAMVKT 517

```

```

; LENGTH: 544
; TYPE: PRT
; ORGANISM: Human
; US-09-422-869-4

Query Match 74.0%; Score 2701.5; DB 4; Length 544;
Best Local Similarity 95.4%; Pred. No. 5.5e-263;
Matches 498; Conservative 5; Mismatches 6; Indels 13; Gaps 1;

QY 1 MRAGGATPARELFRDAAFPAAADSLFCDSLTPLAOFREDITWRRPOICATPRLFPDDP 60
Db 1 MRAGGATPARELFRDAAFPAAADSLFCDSLTPLAOFREDITWRRPOICATPRLFPDDP 60
QY 61 REGOVKOGLLGDCWFLCACAALQKSRHLDDQVIPPQPSMADEYRGSFTCRIMQFGRWV 120
Db 61 REGOVKOGLLGDCWFLCACAALQKSRHLDDQVIPPQPSMADEYRGSFTCRIMQFGRWV 120
QY 121 EYTTDDRLPCLAGRLCFSCQREDVEFWLPLEKYAFAVHGSYEHLMAGVADALVDTLTCG 180
Db 121 EYTTDDRLPCLAGRLCFSCQREDVEFWLPLEKYAFAVHGSYEHLMAGVADALVDTLTCG 180
QY 181 LAERMLKGVASGGGQODRGREHRTCRQLHLKQCLISCCVLSFRAGARELGEFHAF 240
Db 181 LAERMLKGVASGGGQODRGREHRTCRQLHLKQCLISCCVLSFRAGARELGEFHAF 240
QY 241 IYSDLELQAGOCILLRIQNPWGRRCMOGLMREGEGMSQVDAVAASELSQLQEGE 300
Db 241 IYSDLELQAGOCILLRIQNPWGRRCMOGLMREGEGMSQVDAVAASELSQLQEGE 300
QY 301 FWEEBEFLREFDELTVGYPVTEAGHLSYTERLLCHTRALPGAWKQGSAGCRRNSG 360
Db 301 FWEEBEFLREFDELTVGYPVTEAGHLSYTERLLCHTRALPGAWKQGSAGCRRNSG 360
QY 361 FPSNPKFMLRVSPSEYITAVLQSRHLHADNAGARALVGDSTMSPASITGKHYQAV 420
Db 361 FPSNPKFMLRVSPSEYITAVLQSRHLHADNAGARALVGDSTMSPASITGKHYQAV 420
QY 421 GLHLMVEKRRVNLPRVLSMPVAGTACHAYDREVHLRCELSPGYTLAVPSTFLKDAPE 480
Db 421 GLHLMVEKRRVNLPRVLSMPVAGTACHAYDREVHLRCELSPGYTLAVPSTFLKDAPE 480
QY 481 FLIRVSTGRVSLRAIRAVAKTTPGALPAGEMGTVOLGSMRWGOTAGSRNFPASIPT 540
Db 481 FLIRVSTGRVSLRAIRAVAKTTPGALPAGEMGTVOLGSMRWGOTAGSRNFPASIPT 540

```

```

RESULT 5
US-09-422-869-8
; Sequence 8, Application US/09422869
; Patent No. 6235481
; GENERAL INFORMATION:
; APPLICANT: POLONSKY, KENNETH S.
; APPLICANT: HORIKAWA, YUKIO
; APPLICANT: ODA, NAOHISA
; APPLICANT: COX, NANCY J.
; APPLICANT: SREENAN, SEAMUS
; APPLICANT: ZHOU, YUN-PING
; APPLICANT: OTANI, KENICHI
; APPLICANT: HANIS, CRAIG I.
; APPLICANT: BELL, GRAEME I.
; FILE REFERENCE: ARCD:307
; CURRENT FILING DATE: 1999-10-21
; EARLIER APPLICATION NUMBER: US/09/422,869
; EARLIER FILING DATE: 1999-05-13
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 513
; TYPE: PRT
; ORGANISM: Human

```

US-09-422-869-8

Query Match 74.0%; Score 2699; DB 4; Length 513;
Best Local Similarity 98.2%; Pred. No. 8.9e-263;
Matches 496; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

Qy 1 MRAGRGATPARELFRDAAPFAADSSLCFCLSTPLAQFREDITWRRPQEICATPRLFPDDP 60
Db 1 MRAGRGATPARELFRDAAPFAADSSLCFCLSTPLAQFREDITWRRPQEICATPRLFPDDP 60

Qy 61 REGOVKGLLGDCWFLCACAALQSRHLLDQVIPPQPSWADQYRGSTCRIMQFGRW 120
Db 61 REGOVKGLLGDCWFLCACAALQSRHLLDQVIPPQPSWADQYRGSTCRIMQFGRW 120

Qy 121 EVTTDDRLPCLAGRLCFSRCQREDVFWLPLEKYYAKVHGSYEHLMWAGOVADALVLTGG 180
Db 121 EVTTDDRLPCLAGRLCFSRCQREDVFWLPLEKYYAKVHGSYEHLMWAGOVADALVLTGG 180

Qy 181 LAERNLKVAGSGGGQDRPGRWEHRTCRQLLHLKDQCLISCCVLSPRAGARELGEFHF 240
Db 181 LAERNLKVAGSGGGQDRPGRWEHRTCRQLLHLKDQCLISCCVLSPRAGARELGEFHF 240

Qy 241 IVSDRELQOAGOCILLRLIQRNPGRCWQGLWREGGEGHSQYDAVASELLSLOE 300
Db 241 IVSDRELQOAGOCILLRLIQRNPGRCWQGLWREGGEGHSQYDAVASELLSLOE 300

Qy 301 FWEEEFLEFDELTGYPVTEAGHLQSLYTERLCHTRALPGAWKVGOSAGCRNNG 360
Db 301 FWEEEFLEFDELTGYPVTEAGHLQSLYTERLCHTRALPGAWKVGOSAGCRNNG 360

Qy 361 FPSNPKFWLRVSESEVYIAVLRSLRHAADWAGRARALVGDSTSWSPASIPGKHQY 420
Db 361 FPSNPKFWLRVSESEVYIAVLRSLRHAADWAGRARALVGDSTSWSPASIPGKHQY 420

Qy 421 GLHLWK 426
Db 421 GLHLWK 426

Qy 481 FLLRVFSTGRVSLSAIRAVAKNTTP 505
Db 481 FLLRVFSTGRVSLSAIRAVAKNTTP 505

RESULT 6

US-09-422-869-10
; Sequence 10, Application US/09422869
; Patent No. 6235481
; GENERAL INFORMATION:
; APPLICANT: POLONSKY, KENNETH S.
; APPLICANT: HORIKAWA, YUKIO
; APPLICANT: ODA, NAOHISA
; APPLICANT: COX, NANCY J.
; APPLICANT: SREENAN, SEAMUS
; APPLICANT: ZHOU, YUN-PING
; APPLICANT: OTANI, KENICHI
; APPLICANT: HANIS, CRAIG L.
; APPLICANT: BELL, GRAEME I.
; TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
; FILE REFERENCE: ARCD:307
; CURRENT APPLICATION NUMBER: US/09/422,869
; CURRENT FILING DATE: 1999-10-21
; EARLIER APPLICATION NUMBER: 60/134,175
; EARLIER FILING DATE: 1999-05-13
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 10
; LENGTH: 444
; TYPE: PRT
; ORGANISM: Human
US-09-422-869-10

Query Match 64.0%; Score 2336; DB 4; Length 444;
Best Local Similarity 99.8%; Pred. No. 2.3e-226;
Matches 425; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MRAGRGATPARELFRDAAPFAADSSLCFCLSTPLAQFREDITWRRPQEICATPRLFPDDP 60
Db 1 MRAGRGATPARELFRDAAPFAADSSLCFCLSTPLAQFREDITWRRPQEICATPRLFPDDP 60

Qy 61 REGOVKGLLGDCWFLCACAALQSRHLLDQVIPPQPSWADQYRGSTCRIMQFGRW 120
Db 61 REGOVKGLLGDCWFLCACAALQSRHLLDQVIPPQPSWADQYRGSTCRIMQFGRW 120

Qy 121 EVTTDDRLPCLAGRLCFSRCQREDVFWLPLEKYYAKVHGSYEHLMWAGOVADALVLTGG 180
Db 121 EVTTDDRLPCLAGRLCFSRCQREDVFWLPLEKYYAKVHGSYEHLMWAGOVADALVLTGG 180

Qy 181 LAERNLKVAGSGGGQDRPGRWEHRTCRQLLHLKDQCLISCCVLSPRAGARELGEFHF 240
Db 181 LAERNLKVAGSGGGQDRPGRWEHRTCRQLLHLKDQCLISCCVLSPRAGARELGEFHF 240

Qy 241 IVSDRELQOAGOCILLRLIQRNPGRCWQGLWREGGEGHSQYDAVASELLSLOE 300
Db 241 IVSDRELQOAGOCILLRLIQRNPGRCWQGLWREGGEGHSQYDAVASELLSLOE 300

Qy 301 FWEEEFLEFDELTGYPVTEAGHLQSLYTERLCHTRALPGAWKVGOSAGCRNNG 360
Db 301 FWEEEFLEFDELTGYPVTEAGHLQSLYTERLCHTRALPGAWKVGOSAGCRNNG 360

Qy 361 FPSNPKFWLRVSESEVYIAVLRSLRHAADWAGRARALVGDSTSWSPASIPGKHQY 420
Db 361 FPSNPKFWLRVSESEVYIAVLRSLRHAADWAGRARALVGDSTSWSPASIPGKHQY 420

Qy 421 GLHLWK 426
Db 421 GLHLWK 426

RESULT 7

US-09-422-869-12
; Sequence 12, Application US/09422869
; Patent No. 6235481
; GENERAL INFORMATION:
; APPLICANT: POLONSKY, KENNETH S.
; APPLICANT: HORIKAWA, YUKIO
; APPLICANT: ODA, NAOHISA
; APPLICANT: COX, NANCY J.
; APPLICANT: SREENAN, SEAMUS
; APPLICANT: ZHOU, YUN-PING
; APPLICANT: OTANI, KENICHI
; APPLICANT: HANIS, CRAIG L.
; APPLICANT: BELL, GRAEME I.
; TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
; FILE REFERENCE: ARCD:307
; CURRENT APPLICATION NUMBER: US/09/422,869
; CURRENT FILING DATE: 1999-10-21
; EARLIER APPLICATION NUMBER: 60/134,175
; EARLIER FILING DATE: 1999-05-13
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 12
; LENGTH: 274
; TYPE: PRT
; ORGANISM: Human
US-09-422-869-12

Query Match 23.5%; Score 858.5; DB 4; Length 274;
Best Local Similarity 81.7%; Pred. No. 4.6e-78;
Matches 165; Conservative 1; Mismatches 13; Indels 23; Gaps 3;

Qy 1 MRAGRGATPARELFRDAAPFAADSSLCFCLSTPLAQFREDITWRRPQEICATPRLFPDDP 60
Db 1 MRAGRGATPARELFRDAAPFAADSSLCFCLSTPLAQFREDITWRRPQEICATPRLFPDDP 60


```

QY 61 REGVKGGLGDCWFLCACAALOKSRHLLDOVIPPQPSNADQYRGSTFCRIMQGRMV 120
DB 61 REGVKGGLGDCWFLCACAALOKSRHLLDOVIPPQPSNADQYRGSTFCRIMQGRMV 120
QY 121 EVTTDRPLCLAGRLCFSCRCQREDVFWLPLEKRYAKVHGSYEHLMAGVADALVDTLGG 180
DB 121 EVTTDRPLCLAGRLCFSCRCQREDVFWLPLEKRYAKVHGSYEHLMAGVADALVDTLGG 180
QY 181 LAERWNLKGYAGSGGQODRGR 202
DB 168 -----GCPGPDPRRGR 179

RESULT 8
US-08-835-099A-1
; Sequence 1, Application US/08835099A
; Patent No. 5874277
; GENERAL INFORMATION:
; APPLICANT: SHINTANI, Yasushi
; APPLICANT: NISHI, Kazuo
; APPLICANT: KAMAMOTO, Tomohiro
; TITLE OF INVENTION: NOVEL PROTEINS, THEIR PRODUCTION
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/835,099A
; FILING DATE: 04-APR-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 083649/1996
; FILING DATE: 05-APR-1996
; APPLICATION NUMBER: 97105508.2
; FILING DATE: 03-APR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Resnick, David S
; REGISTRATION NUMBER: 34,235
; REFERENCE/DOCKET NUMBER: 47342
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6440
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 703 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-835-099A-1

Query Match 17.4%; Score 635; DB 2; Length 703;
Best Local Similarity 32.3%; Pred. No. 6.5e-55;
Matches 164; Conservative 85; Mismatches 179; Indels 80; Gaps 17;
QY 13 LFRDAFPAADSSLFCOLSTPLAOFREDITRRROEICATYRLPPDDPREGVKGGLG 71
DB 45 LKDDPFPACPSALGKIDLG-PGSPQTOGIWKRPTELCPSPOFTVGGATRTDIOGGIG 103
QY 72 DCMFLCACAALOKSRHLLDOVIPPQPSNADQYRGSTFCRIMQGRMVETTDRLCL 131

```

```

DB 104 DCMFLAIAISLTLLNEELLYRVPPDDP--FOENAGTJFHFOFYGMVEVVIDDLPTK 161
QY 132 AGRICFSCRCQREDVFWLPLEKRYAKVHGSYEHLMAGVADALVDTLGGIAERWNLKGYA 191
DB 162 NGQLLFLHSQGNFQWFLALKEKAYAKLNGCTEALAGSTYGEFDFGTGISEFYDLK--- 218
QY 192 GSGGQODRGRWMEHRCROLLHLKDOC--LISCCV-LSPRAG--RELGEFNAFY 242
DB 219 -----KPPANLYQIIRKAL-----CAGSLGCSIDVSSAAKEATITSKIKSHAYSV 266
QY 243 SDRELQAGAGCCILLRIQNPGRRCQGLMBEGGSCQVDANAASELSLOQEGEFW 302
DB 267 TGVEEVNFO-GHPEKILRLNPMGEVEMSGAWSDDAPEWNNHIDPRRKEELDKRVEDGEFW 325
QY 303 VEEEFLEEFDELTVGVPVTEACHLQSLYTERLLCHTRALPGAMVYQSGCGGRNNGFP 362
DB 326 MSLSDFYROPRLKICMLSPDSLSSEVHKMNVLFN---GHMTRGISTGGQN---YP 378
QY 363 ----SNPKFWLRVSESE-----VYIAVLRSLHADAADNAGRARALVDSHTS 406
DB 379 ATWTNPQFRIRLDEVEDQEEISIGRCCVVLGLMQKNR-----RRKRIGQGLS 430
QY 407 WSPASIPGKHYQAVGLHMKVEKR---RVNLPR-VLSMPVAGTACHAYDREYHLRC 459
DB 431 -----IGYAVVOVPKELESHTDAHGRDFLAYSQARSFYVNLREYSGRA 477
QY 460 ELSPGYIYVSPFELDAGEFLRYES 487
DB 478 RLPGETLVVSPTEEPKDEFLRYES 505

RESULT 9
US-09-157-349-1
; Sequence 1, Application US/09157349
; Patent No. 6068990
; GENERAL INFORMATION:
; APPLICANT: SHINTANI, Yasushi
; APPLICANT: NISHI, Kazuo
; APPLICANT: KAMAMOTO, Tomohiro
; TITLE OF INVENTION: NOVEL PROTEINS, THEIR PRODUCTION
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/157,349
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/835,099.
; FILING DATE:
; APPLICATION NUMBER: 97105508.2
; FILING DATE: 03-APR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Resnick, David S
; REGISTRATION NUMBER: 34,235
; REFERENCE/DOCKET NUMBER: 47342
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6440
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:

```

SEQUENCE CHARACTERISTICS:
LENGTH: 703 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-157-349-1

Query Match 17.4%; Score 635; DB 3; Length 703;
Best Local Similarity 32.3%; Pred. No. 6.5e-55;
Matches 164; Conservative 85; Mismatches 179; Indels 80; Gaps 17;

QY 13 LFRDAAPRAADSS-L-FCDLSTPLAQFREDITWRPQECATRLPDDPREGVQKGLG 71
DB 45 LFKDPEFPACPSALGYKDLG-PGSPQTQGIWKRTPELCPSPQFIVGGATRDICQGLG 103
QY 72 DCWFLCAALQKSHLLDQVIPPQPSWADQEGSTCRWQGRVVEVTTDRPLCL 131
DB 104 DCWLLAAIASLTNEELLYRVPRDQ--FOENYAGIFHFQWQYGEVVEVVDRLPTK 161
QY 132 AGRICFSRCQREDVFWLPLEKVKYAKVHGSYHLWAGQVADALVLTGGLAERNLKGVA 191
DB 162 NGQLFLHSEQNEFWSSALLEKAYAKLNGCYENLAGGSTVEGFTGGISEFYDLK--- 218
QY 192 GSGGQDRPGWEHRTCRQLHLKDDC---LISCCV-LSPRAGA-----RELGEFHAFIV 242
DB 219 -----KPPANLYQIIRKAL-----CAGSLGCSIDVSSAAEAITSQKLVKSHAYSV 266
QY 243 SDRLEQAGQOCILLRIQNPWRCWQGLWREGGEGSWQVDAVAASELLSQLEGFEW 302
DB 267 TGVVEVNFQ-GHPEKILRLNPNWGEVSWGAWSDDAPEWNIHIDPRRKEELDKVEDGEFW 325
QY 303 VEEEFLEFREFDELTVGYPVTEAGHLQSLYTERLCHTRALPGAWKVSAGGCRNNSGFP 362
DB 326 MSLSDFEVROFSRLEICNLSPDSLSEEVHKWNLVFN-----GHWTRGCTAGGCQN---YP 378
QY 363 -----SNPKFWLRYSEPE-----VYIAVLQSLRLHAADWAGARALVGDSDHTS 406
DB 379 ATYTNFQKIRLDEVEDDEQESIGECCTVLLGLMOKNR-----RWRKRIGQGMLS 430
QY 407 WSPASIPGKHQYQVGLHLMKVEKR-----RVNLP--VLSMPPVAGTACHAYDREVLRC 459
DB 431 -----IGYAVIQVPKELESHTDAHLGRDFFLAYQPSARTSTYVNLREVSGRA 477
QY 460 ELSPGYLVAVPSTFLKDPAGEFLLRVFS 487
DB 478 RLPPGEYLVVVPSTFEPFKDGEFCLRVFS 505

RESULT 10
US-08-835-099A-2
Sequence 2, Application US/08835099A
Patent No. 5874277
GENERAL INFORMATION:
APPLICANT: SHINTANI, Yasushi
APPLICANT: NISHI, Kazuori
APPLICANT: KAWAMOTO, Tomohiro
TITLE OF INVENTION: NOVEL PROTEINS, THEIR PRODUCTION
TITLE OF INVENTION: AND USE
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/835,099A
FILING DATE: 04-APR-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 083649/1996
FILING DATE: 05-APR-1996
APPLICATION NUMBER: 97105508.2
FILING DATE: 03-APR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Resnick, David S
REGISTRATION NUMBER: 34,235
REFERENCE/DOCKET NUMBER: 47342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 712 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-835-099A-2

Query Match 17.4%; Score 635; DB 2; Length 712;
Best Local Similarity 32.3%; Pred. No. 6.6e-55;
Matches 164; Conservative 85; Mismatches 179; Indels 80; Gaps 17;

QY 13 LFRDAAPRAADSS-L-FCDLSTPLAQFREDITWRPQECATRLPDDPREGVQKGLG 71
DB 54 LFKDPEFPACPSALGYKDLG-PGSPQTQGIWKRTPELCPSPQFIVGGATRDICQGLG 112
QY 72 DCWFLCAALQKSHLLDQVIPPQPSWADQEGSTCRWQGRVVEVTTDRPLCL 131
DB 113 DCWLLAAIASLTNEELLYRVPRDQ--FOENYAGIFHFQWQYGEVVEVVDRLPTK 170
QY 132 AGRICFSRCQREDVFWLPLEKVKYAKVHGSYHLWAGQVADALVLTGGLAERNLKGVA 191
DB 171 NGQLFLHSEQNEFWSSALLEKAYAKLNGCYENLAGGSTVEGFTGGISEFYDLK--- 227
QY 192 GSGGQDRPGWEHRTCRQLHLKDDC---LISCCV-LSPRAGA-----RELGEFHAFIV 242
DB 228 -----KPPANLYQIIRKAL-----CAGSLGCSIDVSSAAEAITSQKLVKSHAYSV 275
QY 243 SDRLEQAGQOCILLRIQNPWRCWQGLWREGGEGSWQVDAVAASELLSQLEGFEW 302
DB 276 TGVVEVNFQ-GHPEKILRLNPNWGEVSWGAWSDDAPEWNIHIDPRRKEELDKVEDGEFW 334
QY 303 VEEEFLEFREFDELTVGYPVTEAGHLQSLYTERLCHTRALPGAWKVSAGGCRNNSGFP 362
DB 335 MSLSDFEVROFSRLEICNLSPDSLSEEVHKWNLVFN-----GHWTRGCTAGGCQN---YP 387
QY 363 -----SNPKFWLRYSEPE-----VYIAVLQSLRLHAADWAGARALVGDSDHTS 406
DB 388 ATYTNFQKIRLDEVEDDEQESIGECCTVLLGLMOKNR-----RWRKRIGQGMLS 439
QY 407 WSPASIPGKHQYQVGLHLMKVEKR-----RVNLP--VLSMPPVAGTACHAYDREVLRC 459
DB 440 -----IGYAVIQVPKELESHTDAHLGRDFFLAYQPSARTSTYVNLREVSGRA 486
QY 460 ELSPGYLVAVPSTFLKDPAGEFLLRVFS 487
DB 487 RLPPGEYLVVVPSTFEPFKDGEFCLRVFS 514

RESULT 11
US-09-157-349-2
Sequence 2, Application US/09157349
Patent No. 6068990
GENERAL INFORMATION:

APPLICANT: SHINTANI, Yasushi
 APPLICANT: NISHI, Kazuo
 APPLICANT: KAWAMOTO, Tomohito
 TITLE OF INVENTION: NOVEL PROTEINS, THEIR PRODUCTION
 TITLE OF INVENTION: AND USE
 NUMBER OF SEQUENCES: 18
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP
 STREET: 130 Water Street
 CITY: Boston
 STATE: MA
 COUNTRY: USA
 ZIP: 02109
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSeq for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/157,349
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/835,099
 FILING DATE:
 APPLICATION NUMBER: 97105508.2
 FILING DATE: 03-APR-1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Resnick, David S.
 REGISTRATION NUMBER: 34,235
 REFERENCE/DOCKET NUMBER: 47342
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 617-523-3400
 TELEFAX: 617-523-6440
 TELEX:
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 712 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: Protein
 US-09-157-349-2

Query Match 17.4%; Score 635; DB 3; Length 712;
 Best Local Similarity 32.3%; Pred. No. 6.6e-55;
 Matches 164; Conservative 85; Mismatches 179; Indels 80; Gaps 17;

13 LFRDAFPADSSL-FCDDSTPLAOFREDTWRPOICATPRLFPDDPREGVOGGLG 71
 54 LKRDPEFPVPSALGKIDG-PCSPOTGIIWKRPTELCPSPQFIYGATRTDICOGLG 112
 72 DCMFLCACAALOKSRHLDOVTPPGOPSWADQYRSGFTCRIMQFGRWVETTDRLPCL 131
 113 DCMFLAALASLTLNELLYLRYVPRDOD--FOENTAGIFHFQWQGEWYVYIDRLPFRK 170
 132 AGRICFSRCORDEVFWPLLEKYAVKHSYEHLMAGVADALVLDLTGGLAERMLKGYA 191
 171 NGDLFLHSEQNEFMSALKEKAYAKLNGCYALAGSGIVEGFEDFTGISEFTYDK-- 227
 192 GSGGQODRGRWHEHTRCQLHLKDDC--LISCCV-LSPRAGA-----RELGEFHAFTV 242
 228 -----RPANLYOIRKAL-----CAGSLGCSIDVSAEAERISQKLYKSHASV 275
 243 SFLRELQAGACGICLLRLIONPWRRCNOGLMREGGEGHSDVDAVASSELLQDGEFEW 302
 276 TGEVAVNQ-GHEPKLIRNPMGEVEMSGAMSDAPNMNHDPRRKEELDKVDEGEFW 334
 303 VEEEFLEPDELIVGYVTEAGHLQSLYTERLCHTRALPGAWKQSGAGCCRNNSGFP 362
 335 MSLSDVFNQFRLTEICNLSPPSLSEEVHAKNVLTFN-----GHWTGRTAGCCN---YP 387

363 -----SNPKFWLVSERSE-----VYIAVLQSRLLHAADWAGARALVDSHTS 406
 388 ATYTNPOKRIKILDEVEDQOEESIGPCTVILGLMOKNR-----HMRKRIQGMIS 439
 407 WSPASIPGRHYQVGLHKKYER-----RVNLP--VLSMPVAGTACHADREYHLK 459
 440 -----IGYAVQPKELSHTDALHGRFLAYOPSATSTYVNLREVSGRA 486
 460 ELSPEYIYAVPSTFLKDAPEELRYFS 487
 487 RLPPGEYLVPSFPEPKDGECLAVFS 514
 RESULT 12
 US-09-422-869-22
 Sequence 22, Application US/09422869
 Patent No. 6235481
 GENERAL INFORMATION:
 APPLICANT: POLONSKY, KENNETH S.
 APPLICANT: HORIKAWA, YUKIO
 APPLICANT: ODA, NAOHISA
 APPLICANT: COX, NANCY J.
 APPLICANT: SREENAN, SEAMUS
 APPLICANT: ZHOU, YUN-PING
 APPLICANT: OTANI, KENICHI
 APPLICANT: HANIS, CRAIG L.
 APPLICANT: BELL, GRAEME I.
 TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
 FILE REFERENCE: ARCD:307
 CURRENT APPLICATION NUMBER: US/09/422,869
 EARLIER FILING DATE: 1999-10-21
 EARLIER APPLICATION NUMBER: 60/134,175
 NUMBER OF SEQ ID NOS: 30
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 22
 LENGTH: 714
 TYPE: PRT
 ORGANISM: Human
 US-09-422-869-22

Query Match 16.9%; Score 618.5; DB 4; Length 714;
 Best Local Similarity 32.3%; Pred. No. 3e-53;
 Matches 166; Conservative 73; Mismatches 198; Indels 77; Gaps 17;

13 LFRDAFPADSSL-FCDDSTPLAOFREDTWRPOICATPRLFPDDPREGVOGGLG 71
 55 LFRDAFPVPSALGKIDG-PNSSKTYGKIMKRPTELSNPQFIYGATRTDICOGLG 113
 72 DCMFLCACAALOKSRHLDOVTPPGOPSWADQYRSGFTCRIMQFGRWVETTDRLPCL 131
 114 DCMFLAALASLTLNELLYLRYVPRDOD--FQNGYAGIFHFQWQGEWYVYIDRLPFRK 171
 132 AGRICFSRCORDEVFWPLLEKYAVKHSYEHLMAGVADALVLDLTGGLAERMLKGYA 191
 172 DKLIVFVHSAEGNEFMSALKEKAYAKLNGCYALAGSGIVEGFEDFTGISEFTYDK-- 231
 192 GSGGQODRGRWHEHTRCQLHLKDDC--LISCCV-LSPRAGA-----RELGEFHAFTV 242
 232 SFLYQ-----LTKALDERGSLGCSIDISSVLDMEATPFRKLVKSHASV 279
 246 RELQAGACGICLLRLIONPWRRCNOGLMREGGEGHSDVDAVASSELLQDGEFEW 305
 280 KOV-NYRGQVSLIRNPMGEVEMSGAMSDAPNMNHDPRRKEELDKVDEGEFW 338
 306 EELFAPEDELIVGYVTEAGHLQSLYTERLCHTRALPGAWKQSGAGCCRNNSGFP 363
 339 RDMREFTRLTEICNLTLPDA--LKSRTIRK--WNTTLLEGTRGRGTAAGCCRN---YPATF 391
 364 --NPKFWLVSERSE-----VYIAVLQSRLLHAADWAGARALVDSHTS 409
 392 WVNPOKRIKILDEDDDDYGDRESGCSFVLAALMOKR-----RRRRF----- 434

QY 410 ASIPCKHYQAVGLHLMKVEKRRVNLPRV-LSMPPVAGTACHAYD-----REVHLRCELS 462
Db 435 ---GRDMETIGVAYEVPPELVGQPAVHLKRDFFLANASRSEQFINLREVSTRFLP 490
QY 463 PGYILAVPSTFLKDPAGEFLRVF---STGRVSL 493
Db 491 PGEYVVVSTFPNKEGDFVLRFFSEKSGAGTVEL 524

RESULT 13
US-09-422-869-24
; Sequence 24, Application US/09422869
; Patent No. 6235481
; GENERAL INFORMATION:
; APPLICANT: POLONSKY, KENNETH S.
; APPLICANT: HORIKAWA, YUKIO
; APPLICANT: ODA, NAOHISA
; APPLICANT: COX, NANCY J.
; APPLICANT: SREENAN, SEAMUS
; APPLICANT: ZHOU, YUN-PING
; APPLICANT: OTANI, KENICHI
; APPLICANT: HANIS, CRAIG L.
; APPLICANT: BELL, GRAEME I.
; TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
; FILE REFERENCE: ARCD:307
; CURRENT APPLICATION NUMBER: US/09/422.869
; CURRENT FILING DATE: 1999-10-21
; EARLIER APPLICATION NUMBER: 60/134.175
; EARLIER FILING DATE: 1999-05-13
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 821
; TYPE: PRT
; ORGANISM: Human
US-09-422-869-24

Query Match 16.9%; Score 618; DB 4; Length 821;
Best Local Similarity 31.4%; Pred. No. 4.3e-53;
Matches 178; Conservative 79; Mismatches 194; Indels 116; Gaps 19;

QY 13 LFRDAAPAAADSSIFCDLSTPLAQFREDITWRRPOEICATPRLPDDPREGQVQGLGD 72
Db 74 LYVPEFPDPTSLFYQKFP1-QF---VWKRPEICENPRFIDGANTRDICQGLGD 128
QY 73 CWFCLACAAALQKSRHLIDQVIPPQPSWADQYRGSEFTRWQGRVVEVTTDDRLPCL 132
Db 129 CWFCLAACTLQHLFLRVIPHQS--FIENYAGIFHFQWRYGEVWVVDICLPTYN 186
QY 133 GRICFSRCQREDVFWLPLEKVKYAKVHGSYEHLMWAGQVADALVDLTGGLA----- 182
Db 187 NOLVETKSNHRNEFWALSLEKAYAKLHGSYEAALGGNTTEAMEDFTGVAEFFEIRDAPS 246
QY 183 -----ERNWLKGA-----GSGQGDQDRGWRHEHTRCROLLHLKQCLISCC 223
Db 247 DMYKIMKATIERGSLMGCSIDGNTMTYGTSPGLNMGELIARVRNN-----DNSLLQDS 302
QY 224 VLSPRAG-----ARELGEFHAFIVSDRLQOAGOCITLLRIONPWG 266
Db 303 DLDPGRSDERTRTIIPVQYETRMACGLVRCHAYSVTGLDEVPK-GEKVLVRLRNPWG 361
QY 267 RRCQGLWREGGWSQVDAVASELLSQL-QEGEFWVEEFLREFDELTVGYPVTEAG 325
Db 362 QVENGSWSDRKQWSEFVDEKARLQHVTEDEGEFWSYEDFIYHFTKLEICNLATA- 420
QY 326 HQSLYTERLICHTRAL-PCGAWKQSGAGCCRNNSGFP-----SNPKFWLRVSE-----P 374
Db 421 ----LQSDKLTQTTVSVNEGRWRCGSGAGCRN---FPDFTWNTNPOYRLKLEEDDDDD 473
QY 375 SEV-----YIAVQSRSLHAADWAGARALVGDSDHTSNWSPASIPCKHYQAVGLHLMKY--- 427

Db 474 SEVICSFLVALMQNR-----RKDRKL-----GASLFTIGFAIYEVPE 512
QY 428 ---EXRRVNLPRVLSMPPVAGTACHAYDREVHLRCELSPGYILAVPSTFLKDPAGEFLR 484
Db 513 MHGNKQHLQKQDFLLNASKARSKTYINMREVSQRFRLPPSEYVIVPSTYEPHQEGEFLR 572
QY 485 VFTSGR-----VLSAIRAVAKNTT 504
Db 573 VFSEKRNLSSEVENTISVDRPVKKKKT 599

RESULT 14
US-09-422-869-27
; Sequence 27, Application US/09422869
; Patent No. 6235481
; GENERAL INFORMATION:
; APPLICANT: POLONSKY, KENNETH S.
; APPLICANT: HORIKAWA, YUKIO
; APPLICANT: ODA, NAOHISA
; APPLICANT: COX, NANCY J.
; APPLICANT: SREENAN, SEAMUS
; APPLICANT: ZHOU, YUN-PING
; APPLICANT: OTANI, KENICHI
; APPLICANT: HANIS, CRAIG L.
; APPLICANT: BELL, GRAEME I.
; TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
; FILE REFERENCE: ARCD:307
; CURRENT APPLICATION NUMBER: US/09/422.869
; CURRENT FILING DATE: 1999-10-21
; EARLIER APPLICATION NUMBER: 60/134.175
; EARLIER FILING DATE: 1999-05-13
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 27
; LENGTH: 703
; TYPE: PRT
; ORGANISM: RAT
US-09-422-869-27

Query Match 16.8%; Score 612; DB 4; Length 703;
Best Local Similarity 31.2%; Pred. No. 1.3e-52;
Matches 155; Conservative 87; Mismatches 197; Indels 58; Gaps 14;

QY 13 LFRDAAPAAADSSIFCDLSTPLAQFREDITWRRPOEICATPRLPDDPREGQVQKGLG 71
Db 45 LFRDPEFPAPCSALGYKDLG-PGSPDQGVIVWKRPELCPNPQFVIGGATRTDIQGG 103
QY 72 DCWFLCACAALQKSRHLIDQVIPPQPSWADQYRGSEFTRWQGRVVEVTTDDRLPCL 131
Db 104 DCWLLAAIASITLNEKLYRVLPRQS--FKDYAGIFHFQWRYGEVVEVVIDRLPTK 161
QY 132 AGRICFSRCQREDVFWLPLEKVKYAKVHGSYEHLMWAGQVADALVDLTGGLAERNLKGVA 191
Db 162 NGQLLFHSEEGNEFWALSLEKAYAKLHGSYEAALVGGSTIEGFEDEFTGGISEFYDLK--- 218
QY 192 GSGQGDQDRGWRHEHTRCROLLHLKQCLISCCV-LSPRAG-----RELGEFHAFIVSD 245
Db 219 -----KPENLYYIYIQA--LKGSLGCSIDVSTAAEATROKLVGHAYSITGV 269
QY 246 RELQOAGOCITLLRIONPWGRWQGLWREGGWSQVDAVASELLSQLQEGEFWVEE 305
Db 270 EENVFH-GRPEKLIRLNPWGEVSCAWSNAPENNYIDPRKEELDRAEDGEFWSF 328
QY 306 EELREFDELTVGYPVTEAGHLQSLYTERLICHTRALPGAWKQSGAGCCRNNSG-FPSN 364
Db 329 SDFLKQYSRLEICNLSPDSLSESEIHKWNLVFN---GRWTRGTAGGCLNPGTYWTN 384
QY 365 PKFWLRVSEPE-----YIAVQSRSLHAADWAGARALVGDSDHTS--HSPA 410
Db 385 PQKIHLDVEDEQEESTSEPCCTVLLGLMOKNR-----RROKRIQOGMLSIGTAVY 436
QY 411 SIPCKHYQAVGLHLMKVEKRRVNLPRVLSMPPVAGTACHAYDREVHLRCELSPGYILAVP 470

Db 437 QIPKELESHTDAHLGR-----DEFILGRPSTCSSTYMANREVSRRRLPGQLYVP 488
 QY 471 STEFLDAPGEFLLRVES 487
 Db 489 STEEPKDDPCLLRVES 505

RESULT 15

US-08-726-525-7

Sequence 7, Application US/08726525

Patent No. 5789181

GENERAL INFORMATION:

APPLICANT: Lin, Lin-Ling

APPLICANT: Graham, James

TITLE OF INVENTION: NOVEL INTERLEUKIN-1 RECEPTOR

TITLE OF INVENTION: INTRACELLULAR LIGAND PROTEINS AND INHIBITORS OF LIGAND

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESSES:

ADDRESSEE: LEGAL AFFAIRS, GENETICS INSTITUTE, INC.

STREET: 87 Cambridgepark Drive

CITY: Cambridge

STATE: MA

COUNTRY: USA

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/726,525

FILING DATE: 07-OCT-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/487,942

FILING DATE: 07-JUN-1995

ATTORNEY/AGENT INFORMATION:

NAME: Brown, Scott A.

REGISTRATION NUMBER: 32,724

REFERENCE/DOCKET NUMBER: G15258

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 498-8224

TELEFAX: (617) 876-5851

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 700 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

US-08-726-525-7

Query Match

Best Local Similarity 16.6%; Score 604; DB 1; Length 700;

Matches 157; Conservative 76; Mismatches 209; Indels 74; Gaps 15;

QY 10 ARELFRDAPFADSSL-FCDLSPILAQFREDITWRPQICATPRLFPPDPRREGOVKOG 68
 Db 42 AGLTFDPPSPALPSALGFEIG-PISSKTRGMKMRPTICADPOFIIGATRTDICOG 100
 QY 69 LLDCCWFLCAALOKSRHLDDVIPPQGSMDQETRGSGFTCRIMQFGRMVEVTTDDRL 128
 Db 101 ALGDCMFLAIALISITLNEELIAYVPLNOS--FOENYAGIFHFQFMOYGEVVEYVDDRL 158
 QY 129 PCLAGRLCSRCQREDVFNPLLEKYAYAKVHSGYEHLMAQOVADALVLTGGLAERWNLK 188
 Db 159 PTYDGLLEFVHSAEGSEFSALEKAYAKINGCYEALSGATGEGFEDFTGIAEWYELK 218
 QY 189 GVAGSGGQDPRGRWHRCTROLHLKDLCLISCVLSPRAGARELGEF-----HAFTV 242

Db 219 -----KPPNLFKTIQKA--LQKSLGCSIDITSADSEAITFOKLYKHAHSV 266
 QY 243 SDLRELQAGOCICILLRIQNPGRMCQGLMREGEGMSOYDAVASELLSQLOEGEWM 302
 Db 267 TGAEEVESN-GSLQKLRIRNPMGEVETGRNNDCCPSNITIDPERERLTRRHDEGEWM 325
 QY 303 VEEEFLEFDELTVGVPTVTEAGHLQSLYTERILCHT-----RALPGAWYQSGAGGR 356
 Db 326 MSFSDFLRHYSL-----EICNLTPTDITSDTYKKMKLITMDGWMRRGSTAGGR 375
 QY 357 N-NSGFPSNPKFWLRYSEPS-----YTIADVLRSLHAADWAGRARALVGSHT 405
 Db 376 NYPNTFMNPPOLYLIKLEEDDEDEDEDESGCTFLVGLQKHR-----RROKMGEDMHT 428
 QY 406 -SMSPASIPGKHVQAVGLHWK---VEKRRVNLPRVLSMPVAVGTACHADREVLHCE 460
 Db 429 IGFGIYVPEELSGQTNHLSKNFPLTRNARERSPTFTNL-----REVLRPK 476
 QY 461 LSPGYTIAVPTFLKADAPGEFLLRVSTGRVSLSAI 496
 Db 477 LPPEGYILVPTFEPNKGDCFCIRVSEKKADYQAV 512

Search completed: September 19, 2002, 14:49:17
 Job time: 26 sec

us-09-768-877-2.sepl9.rag

20

GenCore version 4.5
(c) 1993 - 2000 Com

2152.295 Million cell updates/sec

.....SØEMLGØFLØEVSVMAVMT 672

747574

seq/geneseqp-emb1/AAL1980.DAT.*
seq/geneseqp-emb1/AAL1981.DAT.*
seq/geneseqp-emb1/AAL1982.DAT.*
seq/geneseqp-emb1/AAL1983.DAT.*
seq/geneseqp-emb1/AAL1984.DAT.*
seq/geneseqp-emb1/AAL1985.DAT.*
seq/geneseqp-emb1/AAL1986.DAT.*
seq/geneseqp-emb1/AAL1987.DAT.*
seq/geneseqp-emb1/AAL1988.DAT.*
seq/geneseqp-emb1/AAL1989.DAT.*
seq/geneseqp-emb1/AAL1990.DAT.*
seq/geneseqp-emb1/AAL1991.DAT.*
seq/geneseqp-emb1/AAL1992.DAT.*
seq/geneseqp-emb1/AAL1993.DAT.*
seq/geneseqp-emb1/AAL1994.DAT.*
seq/geneseqp-emb1/AAL1995.DAT.*
seq/geneseqp-emb1/AAL1996.DAT.*
seq/geneseqp-emb1/AAL1997.DAT.*
seq/geneseqp-emb1/AAL1998.DAT.*
seq/geneseqp-emb1/AAL1999.DAT.*
seq/geneseqp-emb1/AAL2000.DAT.*
seq/geneseqp-emb1/AAL2001.DAT.*

ected by chance to have a
of the result being printed,
score distribution.

Description
Human calpain 10a
Human Protease ancase
Human polypeptide
Mouse calpain 10
Human calpain-like
Human calpain 10c
Human calpain 10b
Human calpain 10d
Human calpain 10e
Human protein sequ
Human calpain 10f

Novel	human	protein
Human	ORFX ORF465	
Human	protein	SEQ
Novel	human	protease
Novel	testis-specific	
Novel	human	protease
Novel	human	protease
Human	calpain	homolog
Human	ORFX ORF1544	
Human	calpain.	H
Human	calpain.	H
Calpain	large subunit	
Mouse	truncated	
Mouse	calpain	nCL
Mouse	calpain	nCL
Mouse	calpain	nCL
Rat	calpain 80kDa	
Human	calpain 80kDa	
A	calpain	protein
Human	ORFX ORF952	
Human	CAMP	used to
Human	interleukin	
Human	calpain	nCL
Drosophila	melanoc	
Drosophila	melanoc	
Human	calpain 10g	
Novel	human	diagnos
Novel	human	enzyme
Human	polypeptide	
Mouse	calpain	CAMP
Human	calpain 10h	
Drosophila	melanoc	

0 gene;
diagnosis;

risks affecting a

PT polymorphism in a calpain encoding nucleic acid segment or a
PT protease-encoding nucleic acid segment -
XX Claim 40; Page 217-219; 257pp; English.
PS The present sequence is that of a novel human cysteine protease,
XX designated calpain 10a. The protein is encoded by the major
transcript (see AA27476) of the human calpain 10 gene, CAPN10 (see
AA27475), which has been identified in the NIDDM1 region of
chromosome 2. The CAPN10 gene consists of 15 exons spanning 32 kb.
A complex pattern of alternative splicing generates proteins of
672, 544, 517, 513, 444, 274, 139 and 138 amino acids (see
AA79567-74), designated calpain 10a to 10h, respectively. The
invention concerns the identification of genes responsible for
type 2 diabetes for use in diagnostic and therapeutic applications.
A G-to-A polymorphism in intron 3 of the CAPN10 gene (UCSNP-43)
shows evidence for linkage to type 2 diabetes. Claimed methods for
screening for a propensity for type 2 diabetes mellitus are based
on detection of a polymorphism in a calpain encoding nucleic acid,
especially UCSNP-10 of the CAPN10 gene. Methods are also claimed
for: producing calpain 10 polypeptides using calpain 10-encoding
polynucleotides; regulating diabetes by providing a calpain
polypeptide such as calpain 10a; identifying modulators of calpain
activity using calpain 10 polypeptides; and using these modulators
to treat diabetes, in particular through the regulation of an
insulin secretory response or insulin mediated glucose transport.

XX SQ Sequence 672 AA;

Query Match		100.0%;	Score 3649;	DB 21;	Length 672;
Best Local Similarity		100.0%;	Pred. No. 0;		
Matches 672;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MRACRGATPARELFRDAFPAADSSLCFSLSTPLAOPREDITWRROEICATRLPPDDP	60		
Db	1	mracrgatparelfrdaafpaadsslcfslstplagfeditwrrpgeicatlrlpddp	60		
QY	61	REGQVKGGLGDCWFLCACAALAKSRLLDQVTPPGQPSWADQEIYRGSFTRCWIQFGRWV	120		
Db	61	regqvkglldgcwflcacaalqksrlldqvpipgqpswadqeyrgsftrciwqgrwv	120		
QY	121	EVTTDDRLPLCLAGLCFSRCQREDFWFLPLEKVKYAKVHGSYEHVLWAGQVADALVDLTGG	180		
Db	121	evttddrlplclaglcfsrcqredvfwpllekvayakvghsyehvlwagqvadaldvtgg	180		
QY	181	LAERNLKVAGSGGQDRPGRWEHRTCTROLLKDKQCLISCCVLSFRAGARELGEFHAF	240		
Db	181	laerwnlkgvaggqdrpgrwehrtctqllhkkqclisccvlspragarelgelghaf	240		
QY	241	IVSDLRELQOAGCCILLRIQNPWRRCWQGLWRGEGGWSQVDAVAASELLSQLOEGE	300		
Db	241	ivsdirelqgagccillirigpwnrrcwqglwreggegwsqvdaavasellslqleqe	300		
QY	301	FWVEEFLREFDELTVGYPTVTEAGHLSLYTERLLCHTRALPGAVKQSGAGCCRNNSG	360		
Db	301	fwveeeeflrefdeltvgyptvteaghlslslyterllchtralpavkqsgagccrnns	360		
QY	361	FPNPKFWLRVSEPEYVIYAVLQSRHLHADWAGARALVGDSTWSPASIPKHYQAV	420		
Db	361	fnpkfwlrsvsepeyviavlrsrlhadwagraralvgdstwspsaspikhyqav	420		
QY	421	GLHLWKEKRVNLPRVLSPPVAGTACHAYDREHVLRELSPGYLAVPSTFLKDAPGE	480		
Db	421	glhlwkekrvnlprvlsppvagtachaydrewhlrelspgyylavpstflkdapge	480		
QY	481	FLLRVFTSGRVSLSAIRAVAKNTTPCAALPAGSGWTVQLRGSRVGTAGGSNFASTYT	540		
Db	481	flrvftsgrvslsairavaknttpcaalpagsgwvqlrgsrvgtaggsnfasptyt	540		
QY	541	NPCFSPVPEGPGRVITLHQHCRPSDTEHPHIFQVPEGGRSQDAPPLLLQEPPL	600		
Db	541	npcfsvpegpgrvritlqhchrpsdtehpghifqvpeggrrsqdapplllqepl	600		

QY	601	LSCVPHRYAOEVSRLCLLPAGTYKVPSTYLPDTGEGFTVTTATRIDRPSHSGMLGOF	660
Db	601	lscvphryaogevsrlcllpagtykvpstylpdtgagftvtctatridrpslshsgmlgof	660
QY	661	LQEVSVMAVMKT 672	
Db	661	lqevsvmavmkt 672	

RESULT 2

AAAY70019			
ID	AAAY70019	standard; Protein; 672 AA.	
XX	AAAY70019;		
XX	05-JUN-2000	(first entry)	
DE	Human Protease and associated protein-13 (PPRG-13).		
XX	Protease and associated protein-13; PPRG-13; anti-PPRG antibody;		
KW	diagnosis; treatment; cell proliferative disorder; cancer; cirrhosis;		
KW	arteriosclerosis; atherosclerosis; bursitis; hepatitis; immune disorder;		
KW	AIDS; Addison's disease; adult respiratory distress syndrome; allergy;		
KW	ankylosing spondylitis; amyloidosis; cytostatic; antiarteriosclerotic;		
KW	hepatototropic; antiinflammatory; virucide; antipsoriatic; anti-HIV;		
KW	antiallergic; immunosuppressive; antidiabetic; antianaemic;		
XX	neuroprotective; human.		
OS	Homo sapiens.		
XX	Key	Location/Qualifiers	
FT	Region	13..322	
FT	Region	/note= "Signature sequence of Calpain"	
FT	Region	57..78	
FT	Modified-site	/note= "Signature sequence of Cysteine protease"	
FT	Modified-site	99	
FT	Modified-site	/note= "Potential Phosphorylation site"	
FT	Modified-site	123	
FT	Modified-site	/note= "Potential Phosphorylation site"	
FT	Modified-site	282	
FT	Modified-site	/note= "Potential Phosphorylation site"	
FT	Modified-site	547	
FT	Modified-site	/note= "Potential Phosphorylation site"	
FT	Modified-site	568	
FT	Modified-site	/note= "Potential Phosphorylation site"	
FT	Modified-site	644	
FT	Modified-site	/note= "Potential Phosphorylation site"	
FT	Modified-site	42	
FT	Modified-site	/note= "Potential Phosphorylation site"	
FT	Modified-site	52	
FT	Modified-site	/note= "Potential Phosphorylation site"	
FT	Modified-site	110	
FT	Modified-site	/note= "Potential Phosphorylation site"	
FT	Modified-site	207	
FT	Modified-site	/note= "Potential Phosphorylation site"	
FT	Modified-site	226	
FT	Modified-site	/note= "Potential Phosphorylation site"	
FT	Modified-site	332	
FT	Modified-site	/note= "Potential Phosphorylation site"	
FT	Modified-site	488	
FT	Modified-site	/note= "Potential Phosphorylation site"	
FT	Modified-site	522	
FT	Modified-site	/note= "Potential Phosphorylation site"	
FT	Modified-site	622	
FT	Modified-site	/note= "Potential Phosphorylation site"	
XX	WO200009709-A2.		
PN	24-FEB-2000.		
PD	06-AUG-1999;		
PF	99WO-US17818.		
XX			

PR 10-AUG-1998; 98US-0096114.
PR 11-FEB-1999; 99US-0119768.

PA (INCY-). INCYTE PHARM INC.

PI Bandman O, Hillman JT, Baughn MR, Azimzai Y, Guegler KT;
PI Corley NC, Yue H, Tang YT, Reddy R, Patterson C, Au-Young J;
PI Shih LL, Lu DAM;

XX WPI: 2000-224346/19.

DR N-PSDB; AAK50930.

PT New human proteases, useful for diagnosis, treatment and prevention of
PT cell proliferative disorders such as atherosclerosis -

PS Claim 1; Page 89-90; 114pp: English.

CC The present sequence is that of human protease and associated protein-13
CC (PPRG-13), which is expressed in nervous and reproductive tissues. It
CC is encoded by cDNA identified in incyte clone 1902576 derived
CC from OVARNO07 cDNA library. Anti-PPRG antibodies can be used
CC as therapeutic antagonists, reagents for diagnosis and monitoring
CC diseases and for isolating PPRG. PPRG nucleotide sequence can be used
CC as probe or primer for diagnosis and monitoring of PPRG-related
CC diseases and gene mapping. PPRG can be used in the treatment of cell
CC proliferative disorders like cancer, arteriosclerosis, atherosclerosis,
CC bursts, cirrhosis and hepatitis, and immune disorders like AIDS,
CC Addison's disease, adult respiratory distress syndrome, allergies,
CC ankylosing spondylitis and amyloidosis.

XX Sequence 672 AA;

Query Match 99.9%; Score 3644; DB 21; Length 672;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 671; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MRAGRGATPARELFRDAAPADSSLCFCDLSTPLAOFREDITWRROEICTATPRLPDDP 60
DB 1 mragrgatparelfrdaapadsslfcfcdlstplagfireditwrroelcatprlfpddp 60
QY 61 REGGVKGLGDCWFLCACAALQKSRHLDOVIRPGOPSNADQYRSGFCRIMQFGRWV 120
DB 61 reggvkglgdcwflcacaalqksrhlldovirppgsnadqeyrsfcrimqfgrwv 120
QY 121 EVTTDDRLPCLAGRLCFRCQREDVFWLPLEKVAHVHSGYELHMGVADALVDTLGG 180
DB 121 evttddrlpclagrlcfrcqredvfwlplekvaahvshgyelhmvgvadaldvdtlvg 180
QY 181 LAERWNLKGVAGSGGODRGRHWRHRCROLHLKQCLISCCVLSRAQARELGEHAF 240
DB 181 laerwnlkgvagsggodrgrhwrhrcrolhlkqclisccvlsraqarelgelhaaf 240
QY 241 IVSDRLRLQAGOCILLRLIIONPWGRRCWGLMREGGEGMSOVDAVASELLSQLQGE 300
DB 241 ivsdrlrlqagocillrlionpwgrrcwglmreggegmsovdavasellsqlqge 300
QY 301 FWVEEFELFEPDELTVGYVTAGHLQSLYTERLLCHTRALPGAWKGGSGAGCRNNSG 360
DB 301 fwveeefelfepdeltvgyvtaghlqsllyterllchtralpawkggsagacrnns 360
QY 361 FPNPKFMLRVSEPEYIIVLQSRRLHADNMGARAVYGDSTMSRPSSTGKHQAV 420
DB 361 fpnpkfmlrvsepeyivlqsrhlhadnmgaravdygdstmsrpsstgkhqav 420
QY 421 GLHLAKREKRRVNLPRVLSNPVAGTACHAYDREVLKCELSPGYTLAVSTFLKADPGE 480
DB 421 glhlakrekrrvnlprvlsnppvagtachaydrevlkcelspgytlavstflkadpge 480
QY 481 FLIRVSTGVSAIRAVAKNTTPGALPAGMGTVQLGSMRVGQTAGGSNRFASYPY 540
DB 481 flirvstgvsaairavaknttpgallpagmgtvqlgsmrvvgqtagsnrfasyp 540

QY 541 NPCEFSVPBEGPCPVCAITLHOCBPSDTEHPHIGFHIQVPEGCSQADAPLLQEP 600
DB 541 npcefsvpbegpcpvcaitlhocbpsdtehp h igfhiqvpegcsqadapllqep 600
QY 601 LSCVPHRYAOFESRLCLPAGYKRVVPSSTYLPDEGACTVLTARPIDRPSHSQEMLCOF 660
DB 601 lscvphryaofesrlclpagykrvvpsstylpdegactvltarpidrpsshsqemlcof 660
QY 661 LQEVSVMAVMKKT 672
DB 661 lqevsvmavmkt 672

RESULT 3
AAM93341
ID AAM93341 standard; Protein: 672 AA.

AC AAM93341;

DT 06-NOV-2001 (first entry)

DE Human polypeptide, SEQ ID NO: 2881.

KM Human: full length cDNA; cDNA synthesis; oligo-capping.

OS Homo sapiens.

PN EPI130094-A2.

PD 05-SEP-2001.

PE 07-JUL-2000; 2000EP-0114089.

PR 08-JUL-1999; 99JP-0194486.

PR 11-JAN-2000; 2000JP-0118774.

PR 02-MAY-2000; 2000JP-0183765.

XX (HELI-) HELIX RES INST.

PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Makamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;

DR WPI: 2001-524255/58.

DR N-PSDB; AAK94261.

PT 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -

PS Claim 8; SEQ ID NO 2881; 1380bp + sequence listing; English.

XX The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesized by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is a polypeptide
CC encoded by a full length human cDNA of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.

XX Sequence 672 AA;

Query Match 99.7%; Score 3638; DB 22; Length 672;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 670; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MRAGRGATPARELFRDAAPADSSLCFCDLSTPLAOFREDITWRROEICTATPRLPDDP 60
DB 1 mragrgatparelfrdaapadsslfcfcdlstplagfireditwrroelcatprlfpddp 60

Db 535 npcpfcvpeagpgrprircltqghcrisdsqblhpfghvfcvpadgengdacslllqep1 594
 Oy 601 lscvphryaagvsnlcllpgtkrvvstltpdtegaftvtlatriDRPSIHQEMLGQF 660
 Db 595 lscvphryaagvsnlcllpgtkrvvstltpdtegaftvtlatriDRPSIHQEMLGQF 654
 Oy 661 lQEVSVAVAK 671
 Db 655 lgevsnfawak 665

RESULT 5
 AAE09756
 ID AAE09756 standard; protein; 517 AA.
 AC AAE09756;
 XX
 DT 29-NOV-2001 (first entry)
 XX
 DE Human calpain-like protease, 18036.
 XX
 KW Calpain-like protease 18036; spleen disorder; splenomegaly; lung;
 KW adult respiratory distress syndrome; colon; liver; jaundice; brain;
 KW idiopathic inflammatory bowel disease; cerebrovascular disease;
 KW acute meningitis; Alzheimer's disease; T-cell; transplant rejection;
 KW systemic lupus erythematosus; skin; seborrheic keratosis; leukaemia;
 KW hematopoietic stem cell; heart; myocardial infarction; atherosclerosis;
 KW kidney; acute proliferative glomerulonephritis; urolithiasis; apoptosis;
 KW ischemia; neurodegenerative disease; demyelinating disease; injury; EAE;
 KW experimental allergic encephalomyelitis; multiple sclerosis; spinal cord;
 KW LGMD2A muscular dystrophy; proliferative disorder; cancer; gene therapy;
 KW human; chromosome 2.
 KW
 OS Homo sapiens.
 XX
 XX
 FH Key Location/Qualifiers
 FT 6..11
 FT Modified-site /note= "N-myristoylation site"
 FT 7..329
 FT Domain /note= "Cysteine protease domain"
 FT 13..322
 FT /label= Calpain_domain
 FT 42..44
 FT Modified-site /note= "Protein kinase C phosphorylation site"
 FT 52..54
 FT Modified-site /note= "Protein kinase C phosphorylation site"
 FT 67..78
 FT Active-site /note= "Eukaryotic thiol (cysteine) proteases
 cysteine active site"
 FT 99..102
 FT Modified-site /note= "Casein kinase II phosphorylation site"
 FT 107..112
 FT Modified-site /note= "N-myristoylation site. The specification
 states that residues GATPAR are present in this site.
 However residues GSFCR are present in the sequence
 given in the specification."
 FT 110..112
 FT Modified-site /note= "Protein kinase C phosphorylation site"
 FT 123..126
 FT Modified-site /note= "Casein kinase II phosphorylation site"
 FT 189..194
 FT Modified-site /note= "N-myristoylation site"
 FT 207..209
 FT Modified-site /note= "Protein kinase C phosphorylation site"
 FT 226..228
 FT Modified-site /note= "Protein kinase C phosphorylation site"
 FT 265..268
 FT Modified-site /note= "Amlidation site"
 FT 282..285
 FT Modified-site /note= "Casein kinase II phosphorylation site"
 FT 332..334
 FT Modified-site /note= "Protein kinase C phosphorylation site"
 FT 338..490
 FT Domain

FT /note= "Calpain large subunit domain"
 FT 349..354
 FT Modified-site /note= "N-myristoylation site"
 FT 357..360
 FT Modified-site /note= "N-glycosylation site"
 FT 467..469
 FT Modified-site /note= "Protein kinase C phosphorylation site"
 FT 489..492
 FT Modified-site /note= "Casein kinase II phosphorylation site"
 XX
 PN M0200164919-A2.
 XX
 PD 07-SEP-2001.
 XX
 PF 26-FEB-2001; 2001MO-US06110.
 XX
 PR 28-FEB-2000; 2000US-0185333.
 XX
 PA (MILL-) MILLENNIUM PHARM INC.
 XX
 PI Kapeller-Libermann R;
 XX
 DR WPI: 2001-565507/63.
 XX
 DR N-PDB; AAD13883.
 XX
 PS New calpain-like protease polypeptides and polynucleotides for
 PT diagnosing, treating seven-transmembrane protein/receptor-related
 PT disorders and to identify modulators of therapeutic use -
 XX
 XX Claim 9; Fig 1; 114pp; English.

CC The invention relates to polypeptide and polynucleotide of calpain-like
 CC protease, 18036. The polypeptide and polynucleotide of the invention are
 CC useful for treatment and diagnosis of calpain-like protease-mediated
 CC disorders. These disorders include disorders of spleen (e.g.
 CC splenomegaly), lung (e.g. adult respiratory distress syndrome), colon
 CC (e.g. idiopathic inflammatory bowel disease), liver (e.g. jaundice),
 CC brain (e.g. traumatic brain injury, cerebral oedema, cerebrovascular
 CC disease, acute meningitis, Alzheimer's disease), T-cells (e.g. transplant
 CC rejection, systemic lupus erythematosus), skin (e.g. seborrheic
 CC keratosis), hematopoietic stem cells (e.g. leukaemia), heart (e.g.
 CC myocardial infarction, atherosclerosis), kidney (e.g. acute proliferative
 CC glomerulonephritis, urolithiasis), thymus, breast, testis, uterus,
 CC epididymis, endometrium, prostate, thyroid, skeletal muscle, pancreas,
 CC small intestine, disorders related to reduced platelet number, bone,
 CC ovary, pain and infectious disorders. The molecules of the invention are
 CC also useful for diagnosing and treating disorders associated with
 CC perturbed cellular growth and differentiation, exercise-induced injury
 CC and repair, apoptosis, ischaemia, neurodegenerative diseases,
 CC demyelinating diseases including experimental allergic encephalomyelitis
 CC (EAE) and multiple sclerosis, LGMD2A muscular dystrophy, spinal cord
 CC injury, proliferative and differentiative disorders e.g. cancer and
 CC renal cell death associated with diverse toxicants. The present sequence
 CC is human calpain-like protease, 18036. The 18036 gene is located on
 CC chromosome 2 between D2S140 and D2S238.
 XX
 SO Sequence 517 AA;

Query Match 74.5%; Score 2719.5; DB 22; Length 517;
 Best Local Similarity 76.9%; Pred. No. 2.1e-238;
 Matches 517; Conservative 0; Mismatches 0; Indels 155; Gaps 1;

Oy 1 NRAGGATPARELFRDAAPADSSDFCDLSPRLAQFREDTWRRROECATPRFPDP 60
 Db 1 mragrgatparelfrdaafpaadssdfcdlspqlaqfiredltwrrpgecatprlfpdp 60
 Oy 61 REGQVKGSLGDQWFLCACALQKSRHLDDQVTPGQFSMAQDEYRGSTCRIMQGRNV 120
 Db 61 regqvkqslldgcwflcacaalqkserhlldqvlppgqswadqerygstctciwqfgrvw 120
 Oy 121 EYTTDRCLPCAGRLCFSGQREDVFWLPLEKVTYAKVGSYTHLMAQVADALVDLNGC 180
 |||

Db 121 evttddrlpclagrlcfsrcqredvfwlpillekvaykvhgshyehlwagqvadalvltgg 180
Qy 181 LAERNLKGAVSGGGQDRGRWEHRTROLLHLKDCCLISCCVLSPRAGARELGEFHAF 240
Db 181 laernlkgavsgggqdrpgrwehrtcrqlhlhkdqcliscvlspragarelgfhaf 240
Qy 241 IVSDRLRQAGOCICILLRIQNPWGRWCQGLWREGGEGWSQVDAVAASELLSQLEGE 300
Db 241 ivsdrlrqagggcicillriqnpwgrwcgllwreggegwsqvaavaseillsqleqe 300
Qy 301 FWEEEFRLREFDELTVGYPTVTEAGHLQSLYTERLLCHTRALPGAWKVGOSAGGCRNNSG 360
Db 301 fweeefrlrefdelvtgypvteaghlqsllyterllchtralpawwkvgsaggrcnns 360
Qy 361 FPSNPKFNLRVSESEVYIAVLRQRSLHAADWAGARALVGDSTWSPASIFPKHYQAV 420
Db 361 fpsnpgkflrvsepsvyyiavlqrsrlhaadwagraralgdshtswspasipgkhyqav 420
Qy 421 GLHLWKVEKRRVNLPRVLSMPPVAGTACHAYDREVHLRCELSPGYIYLAVPSTFLKADPGE 480
Db 421 glhlwk ----- 426
Qy 481 FLRLVFTGTVSLSAIRAVAKNTTPGAALPAGEWGTVOLRGSWRVGTAGGSRNFASYPT 540
Db 427 ----- 426
Qy 541 NPCFFSVPEGPGRCVRIITLHQHCRPSDTEFHPHIGFHFQVPEGGRSQDAPPLLQLEPL 600
Db 427 ----- 426
Qy 601 LSCVPHRYAQEVSRLCLLPAGTKVVPSTYLPDTEGAFVTIATRIDRPSIHSEMIQGF 660
Db 446 lscvphryaqevarcllcpagtkvvpstypdpdegafvtiatridrpsihsqemlqgf 505
Qy 661 LQEVSVMAVNMKT 672
Db 506 lqevsvmavnmkt 517
RESULT 6
ID AAY79569
XX AAY79569 standard; Protein; 517 AA.
AC AAY79569;
XX 15-AUG-2000 (first entry)
DT Human calpain 10c.
DE
DE
DE
KW /NIDDM1; non-insulin-dependent diabetes mellitus; CAPN10 gene;
KW calpain 10; calpain 10c; diapain-1; cysteine protease; diagnosis;
KW therapy.
OS Homo sapiens.
XX
XX WO200023603-A2.
PN
PD 27-APR-2000.
XX
XX 21-OCT-1999; 99WO-US24890.
XX
XX 21-OCT-1998; 98US-0105052.
PR 13-MAY-1999; 99US-0134175.
XX
XX (ARCH-) ARCH DEV CORP.
PA
XX Polonsky KS, Horikawa Y, Oda N, Sreenan S, Zhou Y, Otani K;
PI Hanis CL, Bell GT;
XX
XX WPI: 2000-339702/29.
DR N-PSDB; AAA27475, AAA27578.
XX
XX Method for screening for type 2 diabetes mellitus comprises detecting a

PT polymorphism in a calpain encoding nucleic acid segment or a
PT protease-encoding nucleic acid segment -
XX
PS Claim 40; Page 223-224; 257pp; English.
XX
CC The present sequence is that of a novel human cysteine protease,
CC designated calpain 10c. The protein is encoded by a transcript
CC (see AAA27478) of the human calpain 10 gene, CAPN10 (see AAA27475),
CC which has been identified in the NIDDM1 region of chromosome 2.
CC The CAPN10 gene consists of 15 exons spanning 32 kb. A complex
CC pattern of alternative splicing generates proteins of 672, 544,
CC 517, 513, 444, 274, 139 and 138 amino acids (see AAY79567-74),
CC designated calpain 10a to 10h, respectively. Calpain 10c is
CC readily detectable in many tissues, including skeletal muscle and
CC islets. The invention concerns the identification of genes
CC responsible for type 2 diabetes for use in diagnostic and
CC therapeutic applications. A G-to-A polymorphism in intron 3 of the
CC CAPN10 gene (UCSNP-43) shows evidence for linkage to type 2
CC diabetes. Claimed methods for screening for a propensity for type
CC 2 diabetes mellitus are based on detection of a polymorphism in a
CC calpain encoding nucleic acid, especially UCSNP-10 of the CAPN10
CC gene. Methods are also claimed for: producing calpain 10
CC polypeptides using calpain 10-encoding polynucleotides; regulating
CC diabetes by providing a calpain polypeptide such as calpain 10c;
CC identifying modulators of calpain activity using calpain 10
CC polypeptides; and using these modulators to treat diabetes, in
CC particular through the regulation of an insulin secretory response
CC or insulin mediated glucose transport.
XX
SQ Sequence 517 AA;
Query Match 74.4%; Score 2715.5; DB 21; Length 517;
Best Local Similarity 76.8%; Pred. No. 4.8e-238;
Matches 516; Conservative 1; Mismatches 0; Indels 155; Gaps 1;
QY 1 MRAGRGATPARELFRDAAPFAADSSLFCDLSTPLAQFREDITWRROELCATPLRPDDP 60
DB 1 mragrgatparelfrdaafpaadselcfdstplaqfreditwrpqelcatplrfddp 60
QY 61 REGQVKGILLGDCWFLCACAALQKSRHLLDQVIPPQPSWADQYRGSTCRWQGRWV 120
DB 61 regqvkgillgdcwflcacaalqksrhllldqvippqpswadqyrgstcrwqgrwv 120
QY 121 EVTTDDRLPCLAGRLCFSRCQREDVFWLPILLEKVYKAVHGSYEHHLWAGQVADALVLTGG 180
DB 121 evttddrlpclagrlcfsrcqredvfwlpillekvaykvhgshyehlwagqvadalvltgg 180
QY 181 LAERNLKGAVSGGGQDRGRWEHRTROLLHLKDCCLISCCVLSPRAGARELGEFHAF 240
DB 181 laernlkgavsgggqdrpgrwehrtcrqlhlhkdqcliscvlspragarelgfhaf 240
QY 241 IVSDRLRQAGOCICILLRIQNPWGRWCQGLWREGGEGWSQVDAVAASELLSQLEGE 300
DB 241 ivsdrlrqagggcicillriqnpwgrwcgllwreggegwsqvaavaseillsqleqe 300
QY 301 FWEEEFRLREFDELTVGYPTVTEAGHLQSLYTERLLCHTRALPGAWKVGOSAGGCRNNSG 360
DB 301 fweeefrlrefdelvtgypvteaghlqsllyterllchtralpawwkvgsaggrcnns 360
QY 361 FPSNPKFNLRVSESEVYIAVLRQRSLHAADWAGARALVGDSTWSPASIFPKHYQAV 420
DB 361 fpsnpgkflrvsepsvyyiavlqrsrlhaadwagraralgdshtswspasipgkhyqav 420
QY 421 GLHLWKVEKRRVNLPRVLSMPPVAGTACHAYDREVHLRCELSPGYIYLAVPSTFLKADPGE 480
DB 421 glhlwk ----- 426
QY 481 FLRLVFTGTVSLSAIRAVAKNTTPGAALPAGEWGTVOLRGSWRVGTAGGSRNFASYPT 540
DB 427 ----- 426
QY 541 NPCFFSVPEGPGRCVRIITLHQHCRPSDTEFHPHIGFHFQVPEGGRSQDAPPLLQLEPL 600

```

Db 427 -----|||
601 LSCVPHRYAOEVSRLCLPAGCTKYVPSTYLPRTGCAFTVTATRTDRPSHSOENMGOP 660
446 LSCVPHRYAOEVSRLCLPAGCTKYVPSTYLPRTGCAFTVTATRTDRPSHSOENMGOP 505
661 LQEVSYMAVMKKT 672
506 LQEVSYMAVMKKT 517

RESULT 7
AA79568
ID AAY79568 standard; Protein: 544 AA.
AC AAY79568;
DE 15-AUG-2000 (first entry)
DE Human calpain 10b.
KW NIDDM1; non-insulin-dependent diabetes mellitus; CAPN10 gene;
KW calpain 10; calpain 10b; diapa1n-1; cysteine protease; diagnosis;
KW therapy.
OS Homo sapiens.
PN M0200023603-A2.
DE 27-APR-2000.
DE 21-OCT-1999; 99WO-US24890.
DE 21-OCT-1998; 98US-0105052.
DE 13-MAY-1999; 99US-0134175.
PA (ARCH-) ARCH DEV CORP.
PI Polonsky KS, Horikawa Y, Oda N, Sreenan S, Zhou Y, Otsu K;
PI Hants CL, Bell GI;
DR WPI: 2000-339702/29.
DR N-PSDB: AAA27475, AAA27477.
PT Method for screening for type 2 diabetes mellitus comprises detecting a
PT polymorphism in a calpain encoding nucleic acid segment or a
PT protease-encoding nucleic acid segment -
PS Claim 40; Page 220-222; 257pp; English.
XX The present sequence is that of a novel human cysteine protease,
XX designated calpain 10b. The protein is encoded by a transcript
XX (see AAA27477) of the human calpain 10 gene, CAPN10 (see AAA27475),
XX which has been identified in the NIDDM1 region of chromosome 2.
XX The CAPN10 gene consists of 15 exons spanning 32 kb. A complex
XX pattern of alternative splicing generates proteins of 672, 544,
XX 517, 513, 444, 274, 139 and 138 amino acids (see AAY79567-74),
XX designated calpain 10a to 10h, respectively. Calpain 10b is
XX readily detectable in many tissues, including skeletal muscle and
XX islets. The invention concerns the identification of genes
XX responsible for type 2 diabetes for use in diagnostic and
XX therapeutic applications. A G-to-A polymorphism in Intron 3 of the
XX CAPN10 gene (UCSNP-43) shows evidence for a propensity for type
XX 2 diabetes mellitus are based on detection of a polymorphism in a
XX calpain encoding nucleic acid, especially UCSNP-10 of the CAPN10
XX gene. Methods are also claimed for: producing calpain 10
XX polypeptides using calpain 10-encoding polynucleotides; regulating
XX diabetes by providing a calpain polypeptide such as calpain 10b;
XX identifying modulators of calpain activity using calpain 10
XX polypeptides; and using these modulators to treat diabetes, in
XX particular through the regulation of an insulin secretory response

```

```

CC or Insulin mediated glucose transport.
XX
SQ Sequence 544 AA.
Query Match 74.0%; Score 2701.5; DB 21; Length 544;
Best Local Similarity 95.4%; Pred. No. 9,6e-237;
Matches 498; Conservative 5; Mismatches 6; Indels 13; Gaps 1;
QY 1 MRAGCATPARELFRDAAPPAADSSILFCDSLPLAOFREDITWRARPQECATPRLPPDP 60
DB 1 mragcatparelfrdaappaadssilfcdslplafreditwrrpqcacatprlfpdp 60
QY 61 REGOVKQGLIGDCWFLCACALQKSRHLLDQYIPGQPSWADQEXRGSTCRITWQGRNV 120
DB 61 regovkqgllgdcwflcacalqksrhlldqy1pgqpswadexrgstcr1twqgrnv 120
QY 121 EYTTDRPLCLAGRLCFRSCORDEVFWLPLEKRYAKVGSYEHIMAGOVADALVDLTGG 180
DB 121 eyttdrplclagrlcfrrscordevfwlplekryakvgsyehimagovadalvdl1tgg 180
QY 181 LAERNMLKGVAGSGGQDRPGRMENRTCRQLHLKDOCLISCVLSPRAGARELSEFFNAF 240
DB 181 laernmlkgvagsggqdrpgrmehrtcrqlhlkdqcliscvlspragarelseffnaf 240
QY 241 IYSDARELQGGAGCCILLRIQNPWGRRCWGLMBEGGSGQVDANAVASSELLSQLQEGE 300
DB 241 iysdarelqggagccillriqnpwgrrcwglmbeggsgqvdaavasse1lsq1qege 300
QY 301 FWVEEEFLREDELTVGVPTVEAGHLQSLYERLLCHTRALPGAMVKGOSAGGRNNSG 360
DB 301 fwveeeflrede1tvgyptveaghlqsl1yerllchtralpgawvkgosaggrnns 360
QY 361 FPSNPKFWLRSPESEVYIAVLQSRNLHADWAGARALVGSHTSWSFASIPGRHYQAV 420
DB 361 fpsnpkfwlrspeseyvla1qsrnlhadwagraralvgshs1wsfaspgrhyqav 420
QY 421 GLHMKVEKRRNULPRVLSMPVAGTACHADREYHLKCELSPGYYLAVSPFLDADGE 480
DB 421 glhmkvekrnulp1rvlsmpvagtachadreylh1kcelspgyyla1vpsfl1dadpge 480
QY 481 FLRVSTGRVLSA1RAVAK-----NTTPGAAL 509
DB 481 flrvstgrvls1ra1vaks1c1st1ag1pvt1ps1psa1 522

RESULT 8
AA79570
ID AAY79570 standard; Protein: 513 AA.
AC AAY79570;
DE 15-AUG-2000 (first entry)
DE Human calpain 10d.
KW NIDDM1; non-insulin-dependent diabetes mellitus; CAPN10 gene;
KW calpain 10; calpain 10d; diapa1n-1; cysteine protease; diagnosis;
KW therapy.
OS Homo sapiens.
PN M0200023603-A2.
DE 27-APR-2000.
DE 21-OCT-1999; 99WO-US24890.
DE 21-OCT-1998; 98US-0105052.
DE 13-MAY-1999; 99US-0134175.
PA (ARCH-) ARCH DEV CORP.
XX

```

PI Polonsky KS, Horikawa Y, Oda N, Sreenan S, Zhou Y, Otani K;
PI Hanis CL, Bell GI;
XX WPI; 2000-339702/29.
DR N-PSDB; AAA27475, AAA27479.
DR
XX
PT Method for screening for type 2 diabetes mellitus comprises detecting a
PT polymorphism in a calpain encoding nucleic acid segment or a
PT protease-encoding nucleic acid segment -
XX
XX
PS Claim 40; Page 225-227; 257pp; English.
XX
CC The present sequence is that of a novel human cysteine protease,
CC designated calpain 10d. The protein is encoded by a transcript
CC (see AAA27479) of the human calpain 10 gene, CAPN10 (see AA27475),
CC which has been identified in the NIDDM1 region of chromosome 2.
CC The CAPN10 gene consists of 15 exons spanning 32 kb. A complex
CC pattern of alternative splicing generates proteins of 672, 544,
CC 517, 513, 444, 274, 139 and 138 amino acids (see AAY79567-74),
CC designated calpain 10a to 10h, respectively. Calpain 10d is
CC expressed only at low levels in the tissues examined. The
CC invention concerns the identification of genes responsible for type
CC 2 diabetes for use in diagnostic and therapeutic applications. A
CC G-to-A polymorphism in intron 3 of the CAPN10 gene (UCSNP-43) shows
CC evidence for linkage to type 2 diabetes. Claimed methods for
CC screening for a propensity for type 2 diabetes mellitus are based
CC on detection of a polymorphism in a calpain encoding nucleic acid,
CC especially UCSNP-10 of the CAPN10 gene. Methods are also claimed
CC for: producing calpain 10 polypeptides using calpain 10-encoding
CC polynucleotides; regulating diabetes by providing a calpain
CC polypeptide such as calpain 10d; identifying modulators of calpain
CC activity using calpain 10 polypeptides; and using these modulators
CC to treat diabetes, in particular through the regulation of an
CC insulin secretory response or insulin mediated glucose transport.
XX
SQ Sequence 513 AA;

Query Match 74.0%; Score 2699; DB 21; Length 513;
Best Local Similarity 98.2%; Pred. No. 1.5e-236;
Matches 496; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 1 MRAGRGATPARELFDAAPFAADSSIFCDLSTPLAQREDITWRPQECATPRLFPDPP 60
Db 1 mragrgatparelfrdaafpaadssifcdlstplagrfeditwrrpqecatprlfpddp 60

QY 61 REGQVQGLIGDCWFLCACAAKSRHLLDQVIPPQPSWADQEXRGSTCRWQGRWV 120
Db 61 regqvqgligdcwflcacaaakshrlldqvipppsqswadqeyrgstcrwqgrwv 120

QY 121 EVTTDRPLCLAGRLCFSCQREDVFWLPLEKRYAKVHGSYEHLMWAGQVADALVDLTGG 180
Db 121 evttddrlpclagrlcfscrqredvfwlplekryakvghsyehlmwagqvadalvdtgg 180

QY 181 LAERNLKGAVSGGQODRPGREHETCRQLLHLKDKQCLISCCVLSPRAGRELGFHAP 240
Db 181 laernlkgavsgggodrpgrwehrtcrqlhlhkdkqliscclvspragarelghaf 240

QY 241 IVSDLRELQAGQACILLRIQNPWGRRCWGLWRGEGGWSQVDAVASSELLSQLEGE 300
Db 241 ivsdhrelqagqacillriqnpwgrrcwglwrgeggsqvdavassellsqleqe 300

QY 301 FWTEEFREDFDELJGVPTVTEAGHLQSLYTERLCHTRALPGAWVKQSGAGCCRNNSG 360
Db 301 fwteefrefdeljgvptvteaghlqsllyterlchtralpawvkqsgagccrnns 360

QY 361 FPSNPFVLRVSPSEVYTAVALORSRLHAADWAGRALVGDSTHSWSPASIPGKHQAV 420
Db 361 fspnpxfvlrvspsevylavlsrslhaadwagralvgdsthspsaspigkhyqav 420

QY 421 GLHLKVKERRRNLPRVLSMPPVAGTACHAYDREHLCRCLSPGYVLAVPSTFLKDAPGE 480
Db 421 glhlkwkerrrnlprvlsmppvagtachaydrehlrcrclspgyvlavpstflkdapge 480

QY 481 FLIRVFSTGRVSLSAIRAVAKNTTP 505
Db 481 flirvfstgrvslsairavaknttp 505

RESULT 9

AY79571
ID AAY79571 standard; Protein: 444 AA.

XX AC AAY79571;

XX DT 15-AUG-2000 (first entry)

XX DE Human calpain 10e.

XX KW NIDDM1; non-insulin-dependent diabetes mellitus; CAPN10 gene;
KW calpain 10; calpain 10e; diapa-in-1; cysteine protease; diagnosis;
KW therapy.

XX OS Homo sapiens.

XX PN WO200023603-A2.

XX PD 27-APR-2000.

XX PF 21-OCT-1999; 99WO-US24890.

XX PR 21-OCT-1998; 98US-0105052.

XX PR 13-MAY-1999; 99US-0134175.

XX PA (ARCH-) ARCH DEV CORP.

XX PI Polonsky KS, Horikawa Y, Oda N, Sreenan S, Zhou Y, Otani K;

XX PI Hanis CL, Bell GI;

XX WPI; 2000-339702/29.

XX N-PSDB; AAA27475, AAA27480.

XX Method for screening for type 2 diabetes mellitus comprises detecting a
XX polymorphism in a calpain encoding nucleic acid segment or a
XX protease-encoding nucleic acid segment -
XX Claim 40; Page 228-229; 257pp; English.

XX The present sequence is that of a novel human cysteine protease,
XX designated calpain 10e. The protein is encoded by a transcript
XX (see AAA27480) of the human calpain 10 gene, CAPN10 (see AA27475),
XX which has been identified in the NIDDM1 region of chromosome 2.
XX The CAPN10 gene consists of 15 exons spanning 32 kb. A complex
XX pattern of alternative splicing generates proteins of 672, 544,
XX 517, 513, 444, 274, 139 and 138 amino acids (see AAY79567-74),
XX designated calpain 10a to 10h, respectively. Calpain 10e is
XX expressed only at low levels in the tissues examined. The
XX invention concerns the identification of genes responsible for type
XX 2 diabetes for use in diagnostic and therapeutic applications. A
XX G-to-A polymorphism in intron 3 of the CAPN10 gene (UCSNP-43) shows
XX evidence for linkage to type 2 diabetes. Claimed methods for
XX screening for a propensity for type 2 diabetes mellitus are based
XX on detection of a polymorphism in a calpain encoding nucleic acid,
XX especially UCSNP-10 of the CAPN10 gene. Methods are also claimed
XX for: producing calpain 10 polypeptides using calpain 10-encoding
XX polynucleotides; regulating diabetes by providing a calpain
XX polypeptide such as calpain 10e; identifying modulators of calpain
XX activity using calpain 10 polypeptides; and using these modulators
XX to treat diabetes, in particular through the regulation of an
XX insulin secretory response or insulin mediated glucose transport.

XX Sequence 444 AA;

Query Match 64.0%; Score 2336; DB 21; Length 444;
Best Local Similarity 99.8%; Pred. No. 1.2e-203;

Matches 425; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 MRAGCATPARELFRAAPADSSLFCDLSPLAOFREDIMRRROELCAPRLEPPDP 60
DB 1 mragcatparelfrdaaipaasdlfcdlspqlaqfrdeltrwrgelcaprllppdp 60
OY 61 REGQVKGGLIGDCWFLCACALQKSRHLLDOVIPPQPSMADQETRGSTFCRIMQGRMV 120
DB 61 regqvkqgllgdcwflcacaalqksrhlldqvippqpswadqeyrgstfcrlwqgrmv 120
OY 121 EYTTDDRLPCLAGRLCFSCRCQREDFVWLPLEKRYAVHGSYEHLNAGOVADALVDLNG 180
DB 121 eyttddrlpcclagrlcfscrcqredefvwlplekryavhgsyehlnagovadalvdlng 180
OY 181 LAERNMLKCVAGSGGQODRGRMEHRTCRQLHLKQOCLISCCVLSPRAGARELGEFNAF 240
DB 181 laernmlkcvagsggqodrgmehrtcrqlhlkqocliscvlspragarelgelfnaf 240
OY 241 IVSDRLRELQAGQOCILLRIQNPWGRRCMOGLMREGSGMSQVDAVAASELSQLEGE 300
DB 241 ivsdrrelqagqocillrlqnpwgrrcmoglmregsgmsqvdaavaaselsglqgege 300
OY 301 FVWESEERFREDLTVGTPVTEAGHLSLYTERLLCHTRALPGAWVKQSGAGGCRNNSG 360
DB 301 fvweeseefredlvtvgtpvteaghlsglyterllchtralpawvkgsgaggrnns 360
OY 361 FPSNRFMLRVSESEVYAVQBSRIHADNAGRRALVGSHTSWSASIPGKHQYAV 420
DB 361 fpsnrfmlrvseseyavqbsrihadnagrralvgshtswwsasipgkhyav 420
OY 421 GHLMLK 426
DB 421 ghlmk 426

```

RESULT 10

AAM25421
ID AAM25421 standard; Protein: 245 AA.

AC AAM25421;

DT 16-OCT-2001 (first entry)

DE Human protein sequence SEQ ID NO:936.

XX Human: cancer; ulcer; HIV infection; human immunodeficiency virus;
 KW antiinflammatory; antirheumatic; antiallergic; immunosuppressive;
 KW antibacterial; endocrine; cardiant; central nervous system; vitinide;
 KW anti-HIV; fungicide; antimitagen; cardiovascular; antianemic; anaemia;
 KW antiagregant; haemostatic; vulnerary; antituber; osteopathic; eczema;
 KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;
 KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
 KW immunostimulant; gene therapy; antitense therapy; vaccine; inflammation;
 KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
 KW cardiac dysfunction; neuropathologic disorder; cardiac anaphylaxis; autoimmunity;
 KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
 KW thrombocytopenia; osteoporosis; severe combined immunodeficiency;
 KW allergic rhinitis; diabetes; multiple sclerosis; depression;
 KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
 KW neurological disorder.

XX Homo sapiens.

XX OS AAM25421 standard; Protein: 245 AA.

XX PD 26-JUL-2001.

XX PF 22-DEC-2000; 2000WO-US35017.

XX PR 23-DEC-1999; 99US-0471275.

XX PR 21-JAN-2000; 2000US-0488725.

XX PR 25-APR-2000; 2000US-0552317.

PA (HYSE-) HYSEQ INC.

XX Tang YF, Liu C, Dmanan RT;

XX WPI; 2001-457603/49.

DR N-PSDB; AAM99362.

PT Isolated human polynucleotides encoding polypeptides, useful for the treatment and diagnosis of e.g. cancer, ulcers and HIV infection -

XX Claim 20; Page 201; 1217p; English.

CC AAM99166 to AAM99904 encode the human proteins given in AAM25225 to
 CC AAM25963. The proteins can have activities based on the tissues and
 CC cells they are expressed in, such as: antiinflammatory; antirheumatic;
 CC antiallergic; immunosuppressive; antibacterial; endocrine; cardiant;
 CC central nervous system; vitinide; anti-HIV; fungicide; antimitagen;
 CC cardiovascular; antianemic; antiagregant; haemostatic; vulnerary;
 CC antituber; osteopathic; dermatological; antiallergic; antiasthmatic;
 CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
 CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
 CC encoding them can be used in gene therapy, antisense therapy and vaccine
 CC production. The proteins and polynucleotides are useful for screening for
 CC agonists or antagonists of a protein and for the treatment and diagnosis
 CC of disorders associated with the activity of a protein e.g. inflammation,
 CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
 CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
 CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
 CC anaemia, platelet disorders, thrombocytopenia, wounds, burns, ulcers,
 CC osteoporosis, severe combined immunodeficiency, eczema, allergic
 CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
 CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
 CC neurological disorders.

CC Sequence 245 AA;

Query Match 26.4%; Score 964; DB 22; Length 245;

Best Local Similarity 89.7%; Pred. No. 3.2e-79;

Matches 175; Conservative 2; Mismatches 2; Indels 16; Gaps 1;

OY 60 PREGQVKGGLIGDCWFLCACALQKSRHLLDOVIPPQPSMADQETRGSTFCRIMQGRMV 119

DB 1 pregqvkqgllgdcwflcacaalqksrhlldqvippqpswadqeyrgstfcrlwqgrmv 60

OY 120 VEYTTDDRLPCLAGRLCFSCRCQREDFVWLPLEKRYAVHGSYEHLNAGOVADALVDLNG 179

DB 61 veyttddrlpcclagrlcfscrcqredefvwlplekryavhgsyehlnagovadalvdlng 120

OY 180 GLAERNMLKCVAGSGGQODRGRMEHRTCRQLHLKQOCLISCCVLSPRAGARELGEFNAF 239

DB 121 glaernmlkcvagsggqodrgmehrtcrqlhlkqocliscvlspragarelgelfnaf 171

OY 240 FIVSDRLRELQAGQOCILLRIQNPWGRRCMOGLMREGSGMSQVDAVAASELSQLEGE 234

DB 172 fivsdrlrelqagqocillrlqnpwgrrcmoglmregsgmsqvdaavaaselsglqgege 179

RESULT 11

AAV79572
ID AAV79572 standard; Protein: 274 AA.

AC AAV79572;

DT 15-AUG-2000 (first entry)

DE Human calpain 10f.

XX NIDDM1; non-insulin-dependent diabetes mellitus; CAPN10 gene;

XX calpain 10; calpain 10f; diaphanin-1; cysteine protease; diagnosis;

XX Homo sapiens.

us-09-768-877-2.sepl9.rag

Page 11

antiproliferative; antiparasitism; antineoplastic; neuroprotective;
osteoporosis; osteopenia; osteoarthritis; immunosuppressive;
immunomodulatory; cardiomyopathy; thrombotic; coagulant; vasotonic;
antiinfective; hypotensive; dermatological; immunosuppressive;
antifungal; antibacterial; antiviral; antifungal; antineoplastic;
anthelmintic; and anti-inflammatory. The sequences can be used for determining
the presence of or predisposition to, or preventing or treating
pathological conditions associated with an ORF-associated disorder. The
nucleotide acids can be used to express ORF proteins in gene therapy
vectors. The proteins and nucleotide acids may be used to treat cancers,
proliferative disorders, neurodegenerative disorders, osteoarthritis,
graft vs host disease, cardiovascular disease, diabetes mellitus,
hyperuricaemia, hypothyroidism, cholesterol ester storage, systemic lupus
erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
bacterial or fungal infection, malaria, autoimmune disorders, asthma,
allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
nocturnal haemoglobinuria, antiinflammatory disease, to enhance
coagulation, to inhibit thrombolysis, and as a contraceptive.

SQ Sequence 702 AA;

Query Match	17.78;	Score 644.5;	DB 21;	Length 702;
Post Local string	33.78;	Score 173.40;	DB 10;	Length 302;

Query Match	17.78;	Score 644.5;	DB 21;	Length 702;
Best Local Similarity	33.78;	Pred No 1	70-40.	

matches	100;	conservative	11;	mismatches	201;	indels	59;	gaps	15;
---------	------	--------------	-----	------------	------	--------	-----	------	-----

22 THE ADVANCE F PROSSD F CUSI F LAVE KEDI MWKREVELCAI PRLE PDDPREGV KUGELL /0

OY	71 GDCEFLCACAALQKSRHLLDGVITPPGOSWADOEYRGSEFTCRIMOFGRMWEVYTDDRLPC	130
Dd	gdcwllalalslttcprkillyvvprrgs--fkhyagifhtqvgfwvvvvddrlpt	157
OY	131 LAGRCFSRCDREVFNFMPLLEKYAYAVHGYSYEILMLMGQVALVDYLNGSLAEMNKGV	190
Dd	kndklivfhsiretegfesallekyaklsygealsggstlmeqltedtfgvagvsftqq--	215
OY	191 AGSGGDODREGMEHFTRCOLLHLAKDOCLISCCVLSPBAGARE-----GGEEFAFIIVSD	244
Dd	ipppallrllflfrka--versslmcsleivselsemtkmkvrvghayvtvg	265
OY	245 LRTELQAGAGOTLTLRIONPMGRFCRWOGIMEERGFEGSSOMUAAVAESTIOLIOREFCTE	304

```
Db      266 |advh-vrakmet|rvrnwariwnawdsarawawadlmcgl|htodof.ms 334
```

305 EEEFLREDELTVG--YPTVTEAGHIOSLYTERILCHTRAI.PGAWYKGOSAGCPNNSG-F 361

Db 325 yqdfnnftl|e|c|n|t|p|d|t|s|q|d|y|k|s|w|-----h|t|f|e|q|s|w|r|r|a|s|s|a|a|c|r|h|o|a|t|f 378

QY 362 PSNPKFWLRSE---PSEVYIAVLORSRLHADWAGR---ARALVGDSTHSWSPASIPGK 415

Db 379 wtupgfkislpeddped-----daeqnvvctclvalmqknwrharaqqa 424

QY 416 HYQAVGLHLWKVEKRVNLPV-LSMPVAGTACHAY-----DREVHLRCELSPGYLLA 468

Db 425 qlqtigfvlyavpkfqnigdvhlkfeftkyqdhgfseiftnsrevssqlrlppgeyl 484

QY 469 VPSTFLKDAFGELLRVFS 487

Db 485 ipstfephrdadflrvft 503

RESULT 14

AAM/9025
 ID AAM79025 standard: Protein: 702 aa

AA
AC
AAW79025:

DT 06-NOV-2001 (first entry)

DE Human protein SEQ ID NO 1687.

KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.

OS Homo sapiens.
XX WO200157190-A2.
XX 09-AUG-2001.
XX 05-FEB-2001; 2001WO-US04098.
XX 03-FEB-2000; 2000US-0496914.
XX 27-APR-2000; 2000US-0560875.
XX 20-JUN-2000; 2000US-0598075.
XX 19-JUL-2000; 2000US-0620325.
XX 01-SEP-2000; 2000US-0654936.
XX 15-SEP-2000; 2000US-0663561.
XX 20-OCT-2000; 2000US-0693325.
XX 30-NOV-2000; 2000US-0728422.

PA (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wehrman T, Goodrich R;
XX
DR WPI; 2001-476283/51.
DR N-PSDB; AAK52158.
PT Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
PS Claim 20; Page 4033-4034; 6221pp; English.

XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAK78323-AAK80302) that exhibit activity relating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAK80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.

XX Sequence 702 AA;
Query Match 17.78; Score 644.5; DB 22; Length 702;
Best Local Similarity 33.78; Pred. No. 1.7e-49;
Matches 168; Conservative 71; Mismatches 201; Indels 59; Gaps 15;

Qy 12 ELFRDAFPADSSFL-FCDLSTPLAQFREDITWRPQICATPRLPDPDPREGVKGLL 70
Db 41 elfedplfaeppslgfdlg-psknvqnlswrpkdlnmpifmdgspdtidcgil 99
Qy 71 GDCWFLCACAALQKSRHLLDQVIPPQSPWADQYRGSTCRIFWQFGRVETVTDRLPC 130
Db 100 gdcwllaalgslttcpkilyrvprggs--fkknayagifhfgiqvggwnvvdrrlpt.157
Qy 131 LAGRLCFRCQREDVFWLPLEKVKYAKVHGSVEHLWAGQVADALVDLTGGLAERNLKV 190
Db 158 kndklvfhstersefswallakayaklsyegalsgstmegldftggvafqqlq-- 215
Qy 191 AGSGGQDRPRGWRHRTCRQLHLKDDCLISCCVLSPRAGRE-----LGEFHAFTVSD 244
Db 216 -----rppqnllrlirka--verslmgcsievtstsdselemtkmlrvghaysvtg 265

QY 245 LRELQAGQGCILLRIQNPWRRRCWQGLWREGGWSQVDAVAASELSIQLEGFEFWYE 304
Db 266 lqdvh-ygkmetllirvnpwgricwnagwsdsareweevadigmlhktedgefwns 324
QY 305 EEEFLREFDELTVG--YPTVEAGHLQSLYTERLLCHTRALPGAWKVGQSGGCRNNSG-F 361
Db 325 yqdfllnftlleicnltpdtlsgdyksy-----htfyegswirgssaggrnphgtf 378
QY 362 PSNPKFMLRVSE---PSEVYIAVLQRSRLHRAADWAGR---ARALVCDSTHSWSPASIPCK 415
Db 379 wtnpqfksipegddped-----daegnvvvtclvalmqknvrharqga 424
QY 416 HYQAVGLHLWKVEKRRVNLPRV-LSMPPVAGTACHAY-----DREVHLRCELSPCYYLA 468
Db 425 qlqtigfvlvavpkfqnqdvhlkhefftyqdhgfseiftnsrevsqrllppgeyil 484
QY 469 VPSTFLKADAPGEFLLRVFS 487
Db 485 ipstfephrrdadflirvft 503

RESULT 15
AAB62154
ID AAB62154 standard; Protein; 702 AA.
XX AC AAB62154;
XX 29-MAY-2001 (first entry)
XX Novel human protein (NHP) #3.
XX Novel human protein; NHP; calcium; protease; gene therapy; screening.
XX Homo sapiens.
XX WO200116336-A1.
XX 08-MAR-2001.
XX 01-SEP-2000; 2000WO-US24062.
XX 02-SEP-1999; 99US-0152057.
XX (LEXI-) LEXICON GENETICS INC.
XX Donoho G, Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;
XX WPI; 2001-235112/24.
XX N-PSDB; AAF57304.
XX Novel human polynucleotides isolated from human testis cDNA library,
XX encodes novel human proteins, useful as reagents in screening for
XX compounds used for treating mental and biological disorders -
XX Claim 6; Page 31-33; 39pp; English.

XX The invention relates to novel human polynucleotides encoding protein
XX (NHP) that share sequence similarity with human calcium dependant
XX proteases. Nucleotide constructs encoding functional NHPs are used in
XX gene therapy for the modulation of NHP expression. NHP oligonucleotides
XX can be used as hybridization probes for screening libraries and assessing
XX NHP gene expression patterns. The sequences may also be used as part of
XX ribozyme and/or triple helix sequences that are useful for NHP gene
XX regulation. Labeled NHP nucleotide probes can be used to screen a human
XX genomic library. The NHP nucleotide sequences are also useful in drug
XX screening techniques. Nucleotide constructs encoding NHP products can be
XX used to genetically engineer host cells to express NHP products in vivo,
XX these genetically engineered cells function as bioreactors in the body,
XX delivering a continuous supply of a NHP, a NHP peptide, or a NHP fusion
XX protein to the body. The present sequence represents a NHP.
XX Sequence 702 AA;

Query Match 17.78; Score 644.5; DB 22; Length 702;
 Best Local Similarity 33.78; Pred. No. 1,78-49;
 Matches 168; Conservative 71; Mismatches 201; Indels 59; Caps 15;

```

QY 12 ELFDAAFPADSSL-FCDDSTPLAOFREDITWRPOEICATPRFPDDPREGQKGL 70
Db 41 elifedplfpaeppslgfkdlg-puskvgnslswqrpckdlmnpflfmdgispdalcgdl 99
QY 71 GDCWFLCACAALOKSRHLDOVIRPGOPSWADQETRGSTFCRIWQFGRWVETTDRLPC 130
Db 100 gdcwllaaigslttcpkilyrvprrgs--fknyaglfmftqfvgvwnvvvddrlpt 157
QY 131 LAGRLCSRCQREDFWLPLEKYAKVHGSYEHLMAQVADALVDTLGGIAERMLKGV 190
Db 158 knklivfhstetsetfvsallkayaklsgsyaelsgslmegldftgvgagfqlq-- 215
QY 191 AGSGGQODRPGRWEHFTCRQLHLKQDCLISCCVLSPRAGARE-----LGEFHAFIVSD 244
Db 216 -----ppgnllillika--verslmgsleivtsdelesmtckmlvtrghaysvtg 265
QY 245 LRELQAGAGCCILLRIQNPWGRRCWGLWREGEGMSQYDAAVASELSQLQGEFWE 304
Db 266 lqdvh-yrgkmetllrvnmpwgrlewngawsdaarewevasdldmqllnktdgetkms 324
QY 305 EEEFLAEFDLTVG--YPTAEAGHLSLTYTERLLCHTRALPGAMVKQSNAGCRNNSG-F 361
Db 325 ygdlnnflleicnltptclsgdyksy-----htfyegswirgssagycrnhpftf 378
QY 362 PSNPKFMLRVSE--PSEVYIAVLQSRRLHAADWAGR--ARALYGDHSHTSMPASIPGK 415
Db 379 wtnpgfkislpegddped-----daegnvvvtctivalmqkwntharqga 424
QY 416 HIGAVGLHMKVKRKRVNLEPRV-LSMPVAGTACHAV-----DREVHLRCELSPGYTLA 468
Db 425 qlqltlyfvayvpkefnqldvnlkkekffkyqdhgfseltnsrevsqrlrppgeyl 484
QY 469 VPSTFLKDAPEGELRVES 487
Db 485 ipstcphndadlllrvt 503

```

Search completed: September 19, 2002, 14:49:58
 Job time: 67 sec

